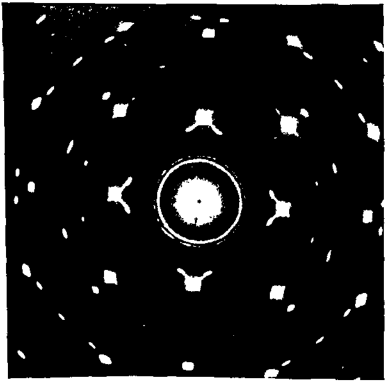
McGraw-Hill Encyclopedia

GRAW-HILL BOOK COMPANY INC.

EW YORK CHICAGO SAN FRANCISCO DALLAS TORONTO LONDON

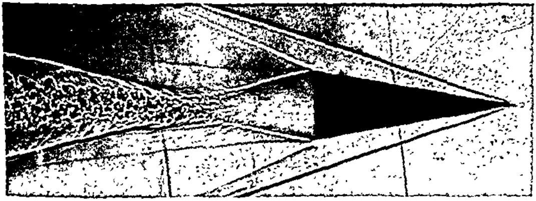


of Science and Technology

AN INTERNATIONAL REFERENCE WORK

IN FIFTEEN VOLUMES INCLUDING AN INDEX

VOLUME 6 GAB-HYS



(LEFT) Electron diffraction pattern from an imperfect gold crystal (*U. S. Bureau of Standards*). (RIGHT) Shadowgraph of a cone in supersonic flight (*National Aeronautics and Space Administration, Ames Research Center*).

Library of Congress Catalog Card Number: 60-11000

Suggestions to the Readers

The basic plan of the Encyclopedia is explained here in order to facilitate its use.

The subject matter of the various disciplines or branches of science and technology is organized systematically: a general article provides a broad survey of the field, and a number of separate articles, alphabetically arranged, cover its main subdivisions and more specific aspects.

Cross references guide the reader from the general articles to the other articles into which the subject is subdivided, and from these to articles on more highly specialized phases of the subject. The cross references—there are about 50,000 of them—are printed in small capital letters so that they can be easily recognized. By means of the cross references a reader may find his way from ELECTRICAL ENGINEERING, through ELECTRONICS and VACUUM TUBE, to ELECTRON MOTION IN VACUUM or ELECTRON EMISSION. Or, following another line of cross references, the reader would be led to ELECTRIC POWER SYSTEMS, TRANSMISSION LINES, ELECTROMAGNETIC WAVE, and so on.

In general, each article begins with a definition of the title that states its scope and coverage. Usually, only the scientific or technological sense is discussed. Most of the articles, after this statement, go on to increasingly complex and detailed considerations. A reader thus needs to proceed only as far as his inclinations and requirements dictate.

The Index, Volume 15, should be consulted to locate the discussion of topics covered in the Encyclopedia but not given in separate entries.

Every phylum, class, and order in the plant and animal kingdoms is allotted a separate article. Many of the more common families, genera, and species are covered either in one of the order articles or in a separate article under its own scientific or common name.

The adjectives electric and electrical are used in the following senses. Electric—containing, producing, arising from, actuated by, or carrying electricity, or capable of doing so; as, for instance, electric generator, electric motor, electric wiring. Electrical—related to, pertaining to, or associated with electricity, but not having its properties or characteristics; as, for example, electrical code, electrical engineering.

Words used as titles are, wherever possible, given in the singular to permit a consistent alphabetic arrangement. Titles are alphabetized by word and not by letter; for example,

Earth sciences
Earth tides
Earthmover
Earthquake

A word used as a noun precedes the same word used adjectivally; thus:

Mercury (element)
Mercury (planet)
Mercury battery

or

Circuit, electronic
Circuit breaker

Hyphened terms are alphabetized as single words; for example,

Animal virus
Animal-feed composition

Most of the longer articles contain bibliographies citing useful sources of further information. For additional bibliographical citations, the reader should refer to related articles (as indicated by the cross references in the article). Bibliographies are placed at the ends of articles or sometimes at the ends of major sections in long articles.

A list of initials and names of the contributors to the Encyclopedia is to be found in Volume 15. This list will permit quick identification of a contributor's initials after an article. Immediately following this list is a second list of encyclopedia contributors with their affiliations and the titles of articles each has written for the Encyclopedia.

McGraw-Hill Encyclopedia of Science and Technology

Gabbro

A phaneritic (visibly crystalline) plutonic rock with granular texture, composed largely of plagioclase (labradorite or bytownite) with slightly smaller amounts of dark-colored (mafic) minerals (pyroxene, olivine, or hornblende). Most gabbros are dark gray or greenish and relatively heavy. For a general discussion of textural, structural, and compositional characteristics *see* Igneous rocks.

Mineralogy. The principal constituent, calcic plagioclase, appears as gray grains of irregular (anhedral) shape or as moderately well-formed (subhedral) tablets. Fine, parallel striations or twin lines may be seen on broken surfaces. Under the microscope, plagioclase uncommonly shows zoning in which more-calcic cores are surrounded by concentric shells progressively more soda-rich. As the composition becomes more sodic than labradorite, the rock passes into diorite.

Mafic minerals may be anhedral or subhedral. Augite is the most common and may contain tiny lamellae of orthopyroxene (exsolved from solid solution). The titanium-rich variety commonly ex- [illegible]

[illegible] gabbros (troctolite) are not abundant. Hornblende is not a common mafic except where it replaces earlier pyroxene. Primary hornblende is usually brown; replacing hornblende is green. Black biotite mica (microscopically brown) is seldom in abundance but may occur with hornblende.

Rocks composed essentially of plagioclase are known as anorthosite. Two types of anorthosite are generally recognized, those with highly calcic plagioclase (labradorite-anorthite) and those with intermediate plagioclase (andesine or labradorite). The latter type transgresses the gabbro-diorite boundary so that technically some representatives are gabbros whereas others are diorites. For simplicity they are all commonly grouped with gabbro.

Quartz and potash feldspar are rarely present in more than trace amounts. They may be intergrown as micropegmatite and are interstitial to other minerals. If quartz exceeds 5%, the rock may be called quartz gabbro. If feldspathoids exceed 5%, the rock is a theralite. Accessories include magnetite, ilmenite, apatite, chromite, and spinel.

Texture. Most gabbros are even grained. Porphyritic texture in which large crystals are sprinkled through the rock is uncommon. Some gabbros show poikilitic texture where large anhedral to subhedral hornblende grains enclose small well-formed (euhedral) feldspar crystals. In other gabbros, pyroxene more or less wraps around euhedral feldspar to give subophitic texture.

Structure. Banding, due to alternation of mafic-rich and mafic-poor layers, is strikingly developed in many gabbros and gives the rock characteristics of sedimentary beds. Bands range from a fraction of an inch to many feet thick and are commonly roughly parallel to the boundaries of gabbro bodies. This relation aids materially in predicting the shape of gabbro masses. The structures appear to have formed as nearly horizontal layers. Highly inclined banding, therefore, suggests disturbance by folding or faulting. Bands may occur in pairs, a light-colored (feldspar-rich) layer gradually passing downward into a mafic-rich layer as if the heavy mafic minerals had settled more quickly than the less dense, light-colored minerals. Such layering is called gravitational banding. Thin gravitational layers may alternate with thicker nonbanded rock to give the well-known rhythmic banding. The cause of these rhythmic structures is not well understood. In addition the features may be accompanied by a flow structure known as igneous lamination in which tabular or platy feldspar and elongate mafic crystals show parallel orientation. Some banded structures may be formed by successive injection of magma of different composition or by the streaking out due to flowage of a heterogeneous magma. According to a more recent concept, the bands may form if crystals develop near the roof of a gabbro body and later settle or become carried down by convection currents to be deposited on the floor. This view is not universally accepted.

Orbicular structures, conspicuous spheroidal aggregates with concentric layers of light and dark minerals, are common. Reaction rims composed of [illegible]

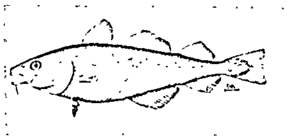
[illegible] magma. It occurs as independent masses (dikes, sills, and stocks) or is associated with quartz diorite, granodiorite, and granite in large batholiths (southern California). Gabbro and norite constitute a major portion of huge layered sheets and bodies (for example, in Transvaal, South Africa; Still-

water complex, Montana; Skaergaard Peninsula, Greenland) where it is associated with considerable mafic-rich rock (peridotite and pyroxenite) and anorthosite.

Anorthosites have two modes of occurrence. The calcic plagioclase type forms thin bands associated with layered gabbros. The intermediate plagioclase type forms gigantic independent bodies such as those of Norway, southeastern Canada, Laramie Range of Wyoming, and the Adirondack Mountains of New York. The intermediate plagioclase type poses a special problem. It may form from anorthositic magma or by differentiation of more mafic (basaltic) magma. Some anorthosites of this class appear to have originated in the solid state by recrystallization and metasomatism of older stratified rocks. *See* MAGMA; METASOMATISM; PETROGRAPHIC PROVINCE. [C.A.CA.]

Gadiformes

A well-defined order of actinopterygian fishes, also known as the Anacanthini, which includes the codfishes and grenadiers. Structurally the group is more or less intermediate between typical soft-rayed and spiny-rayed fishes. As in the former, there are no fin spines, the scales are cycloid, and the pelvic fin often has many rays. As in typical perciforms, however, the swimbladder has no duct, the orbitosphenoid and mesocoracoid are absent, and the upper jaw is bordered only by the protractile premaxillae. The pelvic fins, if present, are jugular in position but are attached to the cleithra only by ligaments. The support of the caudal fin is distinctive in that the terminal vertebra is not upturned and is reduced in size, and the neural and hemal spines from several posterior vertebrae participate in fin support. In the grenadiers the body tapers to a point and the dorsal, caudal, and anal fins are continuous. The opisthotic is very large. Several codfishes have the dorsal divided into three parts, the anal into two.

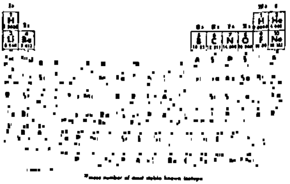


Atlantic cod, Gadus morhua (After G. B. Goode, Great International Fisheries Exhibition, London, 1883, U.S. Natl. Museum Bull. 27, 1884)

Gadiform fishes are known from the Paleocene. Recent forms are classified in 5 families, 63 genera, and nearly 400 species. The families Coryphaenoididae, known as grenadiers or rattails, with nearly 250 species, and Moridae, with about 70 species, include fishes of the deep seas. The best-known family, the Gadidae, includes 60 species that live mostly in shallow to moderate depths of northern seas, where the cods, haddock, pollock, and hakes form the basis for extensive commercial fisheries. One species is widespread in northern fresh waters. *See* ACTINOPTERYGII. [R.M.B.]

Gadolinium

Element number 64, gadolinium, Gd, is a metallic element belonging to the rare-earth group. Its atomic weight is 157.26, and the naturally occurring element is made up of the stable isotopes Gd^{152} 0.200%, Gd^{154} 2.15%, Gd^{155} 14.73%, Gd^{156} 20.47%, Gd^{157} 50.68%, Gd^{158} 24.87%, Gd^{160} 21.90%. In 1880, J. C. G. Marignac obtained a new rare earth in an impure condition and termed it "Y." In 1886, he gave it the name gadolinium in honor of the Swedish scientist J. Gadolin. The oxide, Gd_2O_3, is white, and solutions of the salts are colorless. For properties of the metal, *see* RARE-EARTH ELEMENTS.



Gadolinium salts are of particular interest because they were the first salts used to obtain temperatures below 1°K by means of magnetic cooling. Gadolinium is of particular interest in atomic fission because two of its isotopes, 155 and 157, have extremely large cross sections for thermal neutrons and can act as strong poisons in destroying nuclear chain reactions. For this reason, gadolinium is used to make control rods for nuclear reactors. Gadolinium metal, which is paramagnetic, becomes strongly ferromagnetic below room temperatures. The Curie point occurs at about 16°C. [F.H.SP.]

Gaffkya tetragena

A bacterium, related to the staphylococci, the type species of the genus *Gaffkya*, family Micrococcaceae. The microorganism is highly pathogenic for mice and, while human infections are not common, abscesses of the soft tissues, arthritis, empyema, pneumonia, and meningitis have been reported.

The habitat of this microorganism is similar to that of the staphylococci but it is distinguished from the latter by its occurrence as tetrads, groups of four cocci. The cocci are gram-positive and the colonies are grayish-white and less opaque than those of staphylococci. *See* BACTERIOLOGY, MEDICAL; MICROCOCCACEAE; STAPHYLOCOCCUS. [J.E.B.]

Gages

Devices for determining the relative size or shape of objects. The function of gages is to determine whether parts are within or outside of the specified tolerances, which are expressed in a linear unit of measurement (*see* Tolerance). Gages are the most widely used production tools for controlling linear dimensions during manufacture and for assuring interchangeability of finished parts. A gage may be an indicating type that measures the amount of deviation from a mean or basic dimension, or it may be a fixed type that simply accepts parts within tolerance and rejects parts outside tolerance.

Gage tolerance. If parts are to be interchangeable, it is necessary that, when one device is used to control the accuracy of another, the controlling device itself ordinarily must be accurate to the next significant figure. Thus, if a dimension of a part being manufactured is to be controlled accurately in thousandths of an inch, the gages to do this must be accurate in ten-thousandths of an inch. The gages must be set or measured by instruments accurate in hundred-thousandths of an inch. The instruments must be set by gage blocks accurate in millionths of an inch. The size of the gage blocks is verified by an interferometer reading in ten-millionths of an inch.

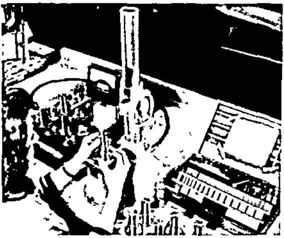


Fig. 2. General purpose comparator of the electrical type is set with gage blocks (in box to right). (*Pratt & Whitney Co.*)

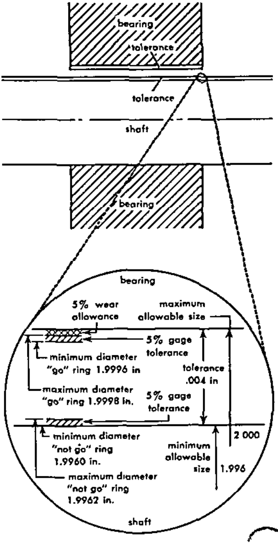


Fig. 1. Make up of gage tolerances.

The tolerance allowed the gagemaker and its effect on the size of the part can be illustrated as follows (Fig. 1). If the diameter of a 2.000-inch shaft is to be controlled during manufacture to 0.004 in. tolerance, the "go" ring gage would have a wear allowance of 0.0002 and a gage maker's tolerance of 0.0002 in. Its size would be 1.9998 − 0.0002 in. The "not go" gage would have no wear allowance, but a gage maker's tolerance of 0.0002 in. Its size would be 1.9960 + 0.0002 in. Thus, because of the necessity of a gage tolerance, if the part tolerance must be maintained absolutely, something less than the full part tolerance is available as a working tolerance, in this case 85%–95%, depending upon the actual sizes of the "go" and "not go" gages.

Gage blocks. The standard of lineal measurement for most manufacturing processes is the gage block (Fig. 2). Gage blocks are used for setting gages, setting machines, and for comparative measurements (*see* Comparator).

Gage blocks are made of chrome steel and may be chrome plated or have carbide wear faces. Individual blocks have two surfaces which are flat and parallel. The parallel distance between the surfaces is the size marked on the block to a guaranteed accuracy of 0.000002, 0.000005, or 0.000008 in. at an ambient temperature of 68°F. Blocks are sold in sets so that combinations of individual blocks, when wrung firmly together, will produce a n end standard equal to the sum of the sizes of individual blocks. With a set of 81 bloc th expressed in four decimal places f

in. to approximately 12.0000 in. can be built up. Between 0.3000 in. and 4.0000 in., 3 to 4 gage groups can be built at the same time from a set to an identical length expressed in four decimal places.

Gage blocks are measured by interferometry. Their actual sizes are determined by comparison with the wavelength of red light in the color spectrum, the ultimate standard of lineal measurement being the wavelength of light (*see* OPTICAL FLAT). Gage blocks translate this basic lineal standard into a practical form for shop use.

Ring gages. Cylindrical rings of steel whose inside diameters are ground and then honed or lapped to gage tolerance are used for checking the diameter of a cylindrical object such as a shaft. The gaging surface is sometimes chrome plated or made of carbide for wear resistance. A single gage controls one limit of tolerance of the diameter. The "go" gage will pass over the largest acceptable part but will not pass over any larger part. The "not go" gage will not pass over any part within tolerance but will pass over an undersize part. A ring gage is superior for checking the "go" dimension because it checks an infinite number of diameters simultaneously and, to the extent of its length, straightness too. Conversely, it is not preferred as a "not go" gage because an out-of-round or bent part could have many diameters undersize, but still be "accepted" because the gage would "not go" if the part were bent or if even one diameter were up to the low limit. Ring gages are made to order.

Plug gage. A cylinder of steel whose diameter is ground, then honed or lapped to gage tolerance, a plug gage is used for checking the diameter of a cylindrical cavity such as a bolt hole or bearing bore. The materials, dimensions, and tolerances of plug gages are comparable to those of ring gages. A single plug gage controls one limit of tolerance of an inside diameter. Plug gages are usually used in pairs, a "go" gage and a "not go" gage. Plug gages are also made to order.

Screw-thread gage. This type of gage may be used to inspect the pitch diameter, major diameter, minor diameter, lead, straightness, and thread angle of a screw thread (Fig. 3). Screw-thread gages may be plug gages, ring gages, snap gages, or indicating gages. A "go" plug or ring gage inspects all these characteristics. A "not go" plug or ring gage checks only pitch diameter and lead. Snap or indicating gages check pitch diameter and one or

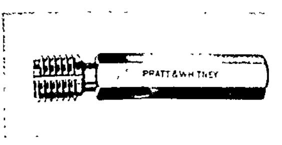


Fig. 3. Thread plug gage. *(Pratt & Whitney Co.)*

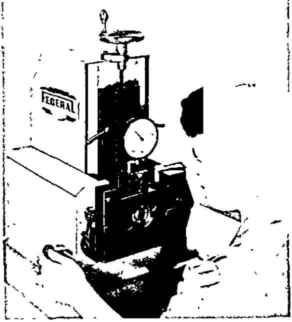


Fig. 4. Indicating gage of the mechanical contact type being used to inspect a part. *(Federal Products Corp.)*

more of the other characteristics. Ring gages are adjustable and can be set with a "setting" plug.

Taper pipe thread gage. Standards for pitch diameter and taper of a pipe thread were first set to control threads used in petroleum service where safety depended on accuracy of the thread dimensions. Standards were set jointly by American Petroleum Institute and U.S. Bureau of Standards. Size is determined by measuring the distance that a plug gage will screw into a coupling or that a ring gage will screw onto a pipe. A manufacturer of pipe or couplings has a master plug and master ring gage certified by an approved laboratory. These are used to control working gages. As gages wear, the distance of engagement changes and limits are reset as long as thread form remains satisfactory.

Snap gage. A snap gage has two surfaces that are flat and parallel, and that are spaced to control one limit of tolerance of an outside diameter or a length. A progressive snap gage has the "not go" surface just behind the "go" surface so that in a single motion both limits of tolerance can be checked. Adjustable snap gages have adjustable gaging surfaces so that, within the limits of frame size, the distance between gaging surfaces can be set with gage blocks and used to gage any dimension falling within these limits.

Snap gaging is faster than ring gaging, but it checks only one diameter or length at a time. It is frequently used as a "not go" gage only, in conjunction with a ring gage for "go" limit to detect out-of-round condition in cylindrical objects. Adjustable gages have an advantage over fixed gages because allowances for wear and gage tolerance may be disregarded in their setting.

Receiving gage. A fixed gage designed to inspect a number of dimensions and also their relation to each other is termed a receiving gage because of its similarity to the cavity which receives the part in actual service. The gage checks on "go" limits. A typical example is a chamber gage which resembles the chamber of a rifle but which is made to the largest limits of size of a cartridge. Thus any cartridge which fits the receiving gage will fit any rifle chamber of that caliber.

Indicating gage. An indicating-type gage has contact points that move as they contact the part being inspected (Fig. 4). The movement is amplified on an indicator whose scale designates the limits of tolerance. Initial setting is done with a master or check gage. An indicator often may be substituted for a fixed or adjustable contact in which a single dimension is being checked. Snap gages are commercially available with indicators instead of fixed contact points. Indicators are also used for inspecting out-of-roundness, centerline runout, or taper, by rotating the part on one axis or diameter and indicating the "run out" of another axis or diameter. Indicating gages do not require allowances for wear or gage tolerance in their setting. *See* INSPECTION AND TESTING.

[R.A.BO.]

Bibliography: F. H. Colvin, *Gages and Their Use in Inspection*, 1942; Pratt and Whitney, *Gages*, 1954; Sheffield Corp., *Gage Laboratory Instruments*, 1957; U.S. Bureau of Standards, *Screw Thread Standards for Federal Services*, 1957.

Gain

A general term used to denote an increase in signal power or voltage produced by an amplifier in transmitting a signal from one point to another. The amount of gain is usually expressed in decibels above a reference level. *See* AMPLIFIER.

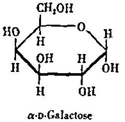
Antenna gain is a measure of the effectiveness of a directional antenna as compared to a nondirectional antenna. It is usually expressed as the ratio in decibels of standard antenna input power to the power input to a directional antenna that will produce the same field strength in the desired direction. The more directional an antenna is, the higher is its gain. *See* ANTENNA (AERIAL). [J.MR.]

Galactose

A monosaccharide and a constituent of oligosaccharides, notably lactose, melibiose, raffinose, and stachyose. It is also known as D-galactose and cerebrose. Agar, gum arabic, mesquite gum, larch arabo galactan, and a variety of other gums and mucilages also contain D-galactose. Crystalline D-galactose has been found in ivy berries.

D-Galactose crystallizes from water as α-D-galactopyranose monohydrate, melting point (mp) 118–120°C, and from hot, anhydrous ethyl alcohol as α-D-galactopyranose, mp 165°C. It shows mutarotation: $[\alpha]_D + 144° \rightarrow 85.5°$ (in water). The anhydrous β-D anomer can be prepared by dissolving α-D-galactose in hot water, cooling to 0°C, and precipitating with alcohol. *See* MONOSACCHARIDE; OPTICAL ACTIVITY.

D-Galactose is not fermentable by bakers' yeast but may be fermented by galactose-adapted and lactose-fermenting yeasts. It enters the main glycolytic pathways by direct interconversion of α-D-galactose-1-phosphate and α-D-glucose-1-phosphate by an enzyme, galactowaldenase, present in the yeast. H. M. Kalckar showed that in the hereditary childhood disease known as galactosemia, which is characterized by abnormal galactose metabolism, the enzyme responsible for the interconversion of the D-galactose-1-phosphate to D-glucose-1-phosphate is greatly lowered in several tissues (erythrocytes, liver). This leads to the accumulation of D-galactose-1-phosphate if D-galactose or lactose is fed.



α-D-Galactose

L-Galactose (enantiomorph of D-galactose) occurs in several polysaccharides including agar, flaxseed mucilage, snail galactogen, and chagual gum. Since D-galactose is usually also present, hydrolysis of these polysaccharides produces DL-galactose. *See* CARBOHYDRATE; POLYSACCHARIDE. [W.Z.H.]

Galaxy, external

A large-scale aggregate of stars, gas, and dust. Galaxies populate the observable portion of the universe with a more or less homogeneous and isotropic distribution. They are separate systems of stars, each containing an average of 10^{11} solar masses and ranging in diameter from 1500 to 300,000 light years. The study of galaxies is divided into two categories: (1) study of the geometrical patterns, stellar content, internal motions, and group properties of individual galaxies with the ultimate aim of finding an evolutionary theory of galactic forms; and (2) study of the distribution of galaxies in space, including their tendency to cluster and to expand and separate.

Individual characteristics. Galaxies are classified in accordance with the appearance of their projected photographic images. The sequence of classification begins with elliptical galaxies (E) which have smooth, elliptical isophotes where the surface brightness falls off as r^{-2} from the bright central region (Fig. 1). The E galaxies range from circular images (E0) to highly flattened forms with a maximum ratio of major-to-minor axes of 3 to 1 (E7).

The classification continues through the spiral forms and then to the irregulars. The spirals are divided into two main groups, the normal spirals (S) with circular symmetry of the nucleus and of

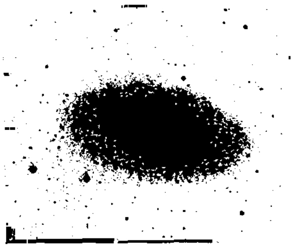


Fig. 1. Elliptical class E5 galaxy NGC 205 in Andromeda photographed in red light by the 200-in. reflector. (*Mount Wilson and Palomar Observatories*)

the spiral arms, and the barred spirals (SB) in which the dominant form is a luminous bar crossing the nucleus with spiral arms starting at the ends of the bar or tangent to a luminous rim on which the bar terminates. Division along the spiral sequence is made according to the openness and degree of resolution of the spiral arms into stars. Types Sa and SBa have smooth, almost circular arms wound tightly about a large, smooth-textured central region. The arms of Sb and SBb galaxies are more open and more irregular than those of Sa and SBa systems. Resolution into stars begins in the arms of Sb and SBb galaxies. The arms of Sc and SBc galaxies are highly resolved into young blue stars, are loosely wound about a small nuclear region, and are highly branched, forming an almost chaotic pattern.

The late spiral (Sd) and irregular forms (Irr) have inconspicuous nuclear regions. The resolved stars, dust clouds, and emission patches dominate the chaotic form.

Stellar content. The composition of galaxies closely parallels the sequence of classification. Normal E galaxies show no obvious dust lanes or clouds nor do they have bright O and B supergiants (*see* STAR). Study of the nearby E galaxies in the local group (NGC 205, M 32, NGC 147, and NGC 185) with the 200-in. telescope shows that the brightest stars are red and are only 600 times the solar luminosity, whereas the brightest stars in the arms of spiral galaxies are 10^5 times brighter than the Sun. The red stars of E galaxies are like the brighter stars in globular clusters in our own galaxy. Theories of evolution show that these red stars are among the oldest stars known (7×10^9 years). The absence of blue stars in E galaxies shows that no new stars have been formed for at least the last 10^9 years and probably longer. Star formation is finished in E galaxies, and we see the remains of stars born far in the past.

On the other hand, the bright resolved blue stars in the arms of spiral galaxies begin to appear in galactic types between Sa and Sb (and SBa and SBb). To be visible with the 200-in. telescope, these stars in the nearest galaxies of the local group [M 31 (Sb), M 33 (Sc), and NGC 6822 (Irr)] have a visible magnitude (M_v) brighter than $M_v = -2$, and many are as bright as $M_v = -9$ (Fig. 2). From the theory of stellar evolution, it is evident that the bright stars are of recent origin, because these stars are consuming their fuel supply at such a rapid rate that they can exist for less than 10^8 years from the time of their formation. The fact that they are seen today shows that these bright stars have been formed recently. Therefore, in contrast to E galaxies, star formation is going on at present in spiral galaxies. There is reason to believe that the formation of spiral arms is connected with this process. *See* STELLAR EVOLUTION.

The stellar content of galaxies in the local group, as studied with large telescopes, consists of all of the brighter types of astronomical objects known in our galaxy (Figs. 3 and 4). The objects which have been identified in the Magellanic Clouds, M 31, M 33, and NGC 6822, include super- [illegible] ulae.

The scale of distances to the galaxies depends on the measurement of the apparent brightness of certain of these objects in particular galaxies. The calibration of the absolute brightness (candlepower) of these distance indicators from data in our own galaxy then gives distances to other galaxies, once the apparent brightness of the same types of objects has been measured.

The stellar content of galaxies which are more distant than the local group and the Virgo cluster cannot be found directly from photographs. However, some information is obtained from the integrated spectrum. Study of such composite spectra reveals the presence or absence of certain types of stars, and the results confirm the difference in stellar content between E galaxies and spirals already

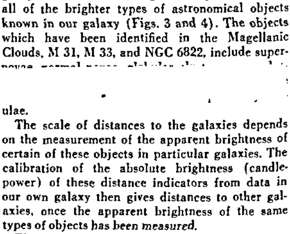


Fig. 2. Central region of class Sc spiral galaxy NGC 598 (Messier 33) in Triangulum; photo by 200-in. reflector. (*Mount Wilson and Palomar Observatories*)

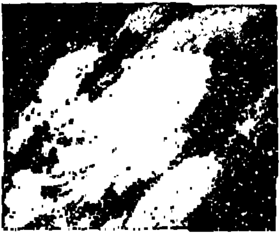


Fig. 3. Central region of Great Spiral NGC 224 (Messier 31) photographed by 100-in. telescope This type Sb flat spiral in Andromeda may be the nearest of the external galaxies. *(Mount Wilson and Palomar Observatories)*

Fig. 4. Great Spiral NGC 224 in Andromeda (Fig. 3) is much like our local galaxy. Above it is elliptical galaxy NGC 205 (Fig. 1), photographed by 48-in. Schmidt telescope. *(Mount Wilson and Palomar Observatories)*

mentioned. Spectra of E galaxies show features due to cool stars, and no component is present due to bright O, B, A, or F stars. Evidence for the presence of F, A, O, and B stars in that order is found as one proceeds along the sequence of classification from the Sa to the Sc galaxy types. All evidence, therefore, shows that the stellar content of galaxies changes along the sequence of classification. This is a vital clue for any evolutionary theory of galaxies.

Geometrical form. Most galaxies are in the form of a flattened disk. The major part of the light of most galaxies comes from this fundamental plane, which is remarkably thin. The degree of flattening is different for the different stellar and gaseous constituents in any given galaxy. For example, most of the gas in our galaxy is concentrated near the fundamental plane with an almost gaussian distribution perpendicular to this plane whose dispersion is about $\sigma = 100$ parsec. This dispersion is small when compared with the diameter of the fundamental plane of more than 20,000 parsec. Most of the intrinsically bright stars such as cepheid variables, O and B stars, and red supergiants show an even flatter distribution with a gaussian dispersion of $\sigma \approx 65$ parsec perpendicular to the plane. Older stars, such as main-sequence A stars and K giant stars (age $\approx 2 \times 10^9$ years), are distributed about the plane with a dispersion of about 250 parsec. On the other hand, long-period variables, RR Lyrae stars, and globular clusters show little or no concentration toward the plane. They form a nearly spherical distribution centered on the galactic nucleus with a diameter of the distribution of at least 20,000 parsec.

To distinguish the distributions, they are termed the nuclear, the disk, the spiral-arm, and the halo populations. The cepheid variable stars and the O and B stars contribute to the spiral-arm population. These stars are all younger than 10^9 years. The halo population, composed of globular cluster-like stars and RR Lyrae variables, contains the oldest stars in the galaxy; these stars are estimated to be 7×10^9 years old. Information on the disk and nuclear populations is still scanty. The separation of stars according to age and distribution is usually considered as a clue to the unsolved problem of galactic evolution.

Internal motions. Within galaxies the stars and associated matter have two motions: (1) a systematic rotation of the entire system about an axis passing through the nucleus and perpendicular to the fundamental plane, and (2) irregular or random motions of the constituent stars and gas clouds —a motion which is superposed on the rotational motion.

Systematic rotation. All galaxies appear to be rotating. The evidence is of two kinds, observation of ionized hydrogen and measurement of spectra. Observations of the radial velocity of H II regions in M 31 and M 33, selected to cover the major axes of the spirals, show that the velocity relative to the center of these galaxies first increases with increasing distance from the center and then decreases as $r^{-1/2}$ at large distances. The rotational velocity of the disk at a distance of 10^4 parsec from the center of M 31 is about 250 km/sec. Spectral lines, both emission and absorption, are inclined on spectrograms taken with the slit along the major axis of galaxies which are seen nearly edge on, that is, with their fundamental plane nearly on the line of sight.

The rate of rotation for most galaxies studied so far is of the order of 3×10^{-8} radian/year. This corresponds to a period of rotation of 2×10^9 years. N. U. Mayall has produced evidence that the rotation period may vary along the sequence of classification but he considers the evidence to be preliminary and subject to confirmation. At present, there is no reliable information on how th

angular momentum per unit mass changes along the classification sequence.

Rotation must be considered as a major feature of galaxies. At first, it might appear that the flattened shape of the galaxies is a result of the rotation. This is further suggested by the close coincidence of the axis of rotation to the axis of symmetry of the galaxies. But rotation can flatten only a distribution of matter which interacts with itself by inelastic collisions. The particles of a fluid or the molecules of a gas form such systems. In the case of galaxies, however, the stars do not interact with one another in times of the order of 10^{10} years, which is the present age of galaxies. The time of relaxation for the stars in a galaxy is long (10^{14} years) compared with the age of galaxies (10^{10} years). This makes necessary the conclusion that the present form of the galaxies is a result of rotation which flattened the matter before the stars were formed at a time when the galaxies were composed of a highly interacting prestellar material such as gas.

Random motion. The internal random motion of the stars in a given galaxy is complex. Although the galaxies rotate as a whole about their centers, the individual stars do not travel in circular orbits but rather move in elliptical paths of varying eccentricities. As seen from a point moving in a circular orbit about the center, the apparent motions of the stars in the neighborhood of that point will appear to be distributed along three perpendicular axes. Study of this distribution permits the calculation of the eccentricity of the orbit of any star. Such studies show that stars of the halo population of our galaxy are moving in highly eccentric orbits about the galactic center with a smaller net rotational motion than stars of the disk population.

In addition to the peculiar motions of the stars, the gas clouds also have random velocities with respect to the rotational velocity. The mean random speed of the gas clouds is estimated at about 10 km/sec as deduced from spectrograms showing absorption lines due to the interstellar gas. The source of energy which must be pumped into the clouds to maintain this random motion against the sources of dissipation is believed to be radiation from the O and B stars usually associated with gas clouds.

Mass of galaxies. An estimate of the mass of a galaxy is obtained from the rate of rotation by considering that the galaxy in question is in equilibrium under the centrifugal force of rotation and the gravitational force of the distribution of the mass. The rotational data for only a few galaxies have been well enough determined to apply the method. The results give a mass of 3×10^{11} solar masses ($M\odot = 6 \times 10^{44}$ g) for M 31, 2×10^{9} $M\odot$ for M 33, 3×10^{9} $M\odot$ for the Large Magellanic Cloud, and 1×10^{9} $M\odot$ for the Small Magellanic Cloud.

The only other method of mass determination is to apply the virial theorem to a cluster of galaxies. This method assumes that there is an equilibrium between the kinetic energy of the random motion, which acts to separate the galaxies, and the gravitational potential energy, which acts to contain the galaxies. Average masses of 10^{11} $M\odot$ are found by this method—a value higher than expected from the mean masses derived from M 31, M 33, and the Magellanic Clouds. There may be objections to the virial theorem method because the aggregates of galaxies to which the equation is applied may not be in equilibrium.

Brightness. Absolute magnitudes of galaxies in optical wavelengths show a wide variation. The best determined values are for galaxies in the local group. The absolute photographic magnitudes (M_{pg}) of typical members of the local group are M 31, $M_{pg} = -20.3$; M 32, $M_{pg} = -15.5$; M 33, $M_{pg} = -17.8$; NGC 185, $M_{pg} = -14.3$; NGC 6822, $M_{pg} = -15.6$. By far the brightest of these galaxies is M 31. There are, however, many galaxies as bright and brighter than M 31 in the general field beyond the local group.

The range in absolute magnitude is from $M_{pg} = -22.5$ to fainter than $M_{pg} = -12.5$, a range which spans a factor of 10^4 in intrinsic candlepower. From statistical data on 1000 galaxies of all types where some estimate of the distance is available, there is no difference in the distribution of absolute magnitude nor in the mean value of M_{pg} between the various types along the sequence of classification, except for the irregular systems which appear to be fainter.

System of galaxies. The distribution of galaxies in space shows a definite tendency of galaxies to cluster. Our own galaxy is at one edge of a small cluster, known as the local group, which contains about 15 members, most of which are of low luminosity. The nearest galaxies beyond the local group are located in two clusters centered on the large spirals M 81 and M 101. Each of these clusters contains about 15 visible members. The faintest members of the local group (the dwarf ellipticals in the constellations of Draco and Ursa Minor) are too faint to be detected if placed at the distance of the M 81 and M 101 groups. The actual number of members of the M 81 and M 101 groups is undoubtedly greater than 15.

Isolated galaxies in the general field are rare. Small groups containing from two to thirty galaxies are numerous. These groups are smaller examples of great clusters such as the Virgo cluster or the Coma cluster, each of which contains several thousand members. If vast volumes of space are surveyed, the clustering tendency of galaxies averages out and the distribution appears to be homogeneous and isotropic. The volume over which the average is taken must of course be large compared with the diameter of a typical cluster for this property of homogeneity to hold.

The average space density of galaxies is of the order of 2×10^{-76} galaxy/cm^3, if a distance scale defined by the Hubble expansion parameter of 75 km/(sec) (10^6 parsec) is used. There is one galaxy for every 5×10^{75} cm^3 of space or one galaxy for

Andromeda Nebula, M 31 or NGC 224, the spiral galaxy which is nearest the Milky Way system Also called the Great Nebula, it rotates about its center with a period of app ly 2×10^8 years and is believed to be some 2,500,000 light years distant It contains b' stars, bright supergiant red stars, ; stars, star clusters, planetary gas, and dust. *Photograph made* *midt telescope* on *Mount Palomar* *Wi'* *and Palomar* Observatories)

every cube of 2×10^7 light years on a side. These numbers, together with an assumed average mass of 2×10^{11} solar masses per galaxy, give the mean density of visible matter in galaxies as 10^{-31} g/cm^3. This should be compared with the mean density of matter in the solar neighborhood at the outskirts of our galaxy of 10^{-24} g/cm^3 and with a density required by cosmological theory of 10^{-29} g/cm^3.

The number of galaxies in a sphere extending to the Hydra cluster, which is near the optical limit of the 60-in. telescope on Mount Wilson ($r = 3 \times 10^{27}$ cm $= 10^9$ parsec) is about 3×10^7. The Hydra cluster is near the spectroscopic limit of the 200-in. telescope. It is estimated that the 200-in. telescope can photograph galaxies which are three to four times more distant than the Hydra cluster. Consequently, the total number of galaxies within the optical range of the 200-in. should be of the order of 10^9.

Expanding universe. It was discovered in 1929 that the universe is expanding; the clue to this discovery was the observation that the spectral features of all galaxies are shifted toward the red. This red shift of the Fraunhofer lines was interpreted as a Doppler effect, which showed that the radial velocities of the galaxies are positive and that the galaxies are receding from the Earth. The measured velocities increase directly with the distance, following the empirical law

$$c\Delta\lambda/\lambda = Hr$$

where $c\Delta\lambda/\lambda$ is the velocity of light multiplied by the fractional wavelength shift of features in the spectrum, H is a proportionality factor, often called the Hubble constant, and r is the distance. This linear velocity-distance relation is known to hold for distances of about 10^9 parsec, which corresponds to apparent velocities of 60,000 km/sec. The apparent expansion of the universe is the most important point of contact of the theory of cosmology with observational astronomy (*see* COSMOLOGY). A linear law is required in all cosmological theories which postulate a form of the Copernican principle that the Earth should not be a favored place in the universe; that is, that the large-scale aspect of the universe will look the same at any arbitrary place as it does from Earth On this point, the observed linear expansion law is in exact agreement with theory.

Red shifts are determined from spectroscopic observations of galaxies made with large telescopes. Measurement of the wavelength displacement from the laboratory wavelengths of identified Fraunhofer lines in spectra of galaxies gives a number $\Delta\lambda$. All tests to date show that $\Delta\lambda$ is proportional to the laboratory wavelength λ_0. Hence, $\Delta\lambda/\lambda_0$ is a constant for a given galaxy. This constancy is required if red shifts are true velocity shifts. There is, therefore, every reason to believe that the observed red shifts represent actual expansion.

The physical dimensions of H are (seconds)$^{-1}$. On certain cosmological theories the value of $1/H$ is related to the time since the beginning of the expansion. On the distance scale of galaxies first derived by Hubble in 1936, H had the value of about 550 km/(sec) (10^6 parsec). Modern estimates, using observations made with the 200-in. telescope, give $H = 75$ km/(sec) (10^6 parsec). This change from Hubble's value represents a correction of a factor of about 7 in the distances to galaxies beyond the local group. The smaller value of H gives a characteristic time, often called the Hubble time, of $H^{-1} = 13 \times 10^9$ years. The cosmological model which gives H^{-1} as the time since the beginning of the expansion is one in which the density of matter is zero and the universe is open, infinite, and has a curved intrinsic geometry of hyperbolic type.

A model, which is more realistic in view of the observational data, is that of an expanding euclidean (flat) space with an infinite radius of curvature. In this case, the time since the beginning of the expansion is $2/3H^{-1} = 9 \times 10^9$ years. For this model, Einstein's field equations for the general theory of relativity require space to have a density of $\rho = 3H^2/8\pi G = 10^{-29}$ g/cm^3.

All cosmological models predict departures from linearity in the observed velocity-distance relation because as one looks out in space one looks back in time and sees the universe as it was a time ago equal to the time it takes light to travel from the distant galaxies to the viewer. Any change in the expansion rate in this interval will show up as a departure from linearity in the observed red-shift–distance relation. Different models predict different results. Present observations of the departures from linearity slightly favor a model whose intrinsic geometry is that of a closed space (finite radius of curvature), whose density at present is about $6H^2/8\pi G = 2 \times 10^{-29}$ g/cm^3, with a present radius of curvature of $c/H = 13 \times 10^9$ light years, with a total mass of 8×10^{56} g, with a total number of galaxies of about 2×10^{13}, and where the expansion will eventually stop and contraction will begin. Because of the potentiality of a decision between models, much observational work has been done since the 200-in. telescope went into operation, to obtain red shifts of galaxies, to estimate their distances, and to determine the departure from linearity in the velocity-distance relation. Initial observations were crude, but improvements in the experimental accuracy together with belief in the theory are expected to give definite answers.

The required cosmological density of $6H^2/8\pi G = 2 \times 10^{-29}$ g/cm^3 is about 100 times greater than can be accounted for by counting the galaxies and assigning a maximum mass of 2×10^{11} solar masses to them. This procedure gives $\rho = 10^{-31}$ g/cm^3. Hence, either the theory is incorrect or about 100 times more matter exists between the galaxies in the form of nonluminous matter than exists in the galaxies. Astronomers are not yet in a position to decide which possibility is correct. *See* INTERSTELLAR MATTER; STAR CLUSTERS; STELLAR EVOLUTION; *see also* ASTRONOMY. [A.R.S.]

Galaxy, the

The flattened spiral system of stars and interstellar material in which our Sun and its solar system are located. The Milky Way, the faint band of light which circles the entire celestial sphere, is the integrated effect of the enormous number of individually invisible stars seen when looking through the plane of the galaxy from our position within it.

With only three exceptions, everything that can be seen with the naked eye belongs to our galaxy. All the exceptions are other galaxies relatively close to our own; they are the two Magellanic Clouds and the Andromeda Galaxy. For a description of these objects, *see* ANDROMEDA NEBULA; MAGELLANIC CLOUDS.

Shape, structure, and dimensions. Our galaxy is a spiral system, as can be deduced by examining external galaxies of different types and in different orientations and comparing their appearance and composition to our own galaxy. The pronounced bulge in the Milky Way in the constellation Sagittarius, particularly conspicuous when photographed in infrared light, which penetrates the interstellar material more readily than visible light, is the nucleus of our galaxy as seen from our vantage point roughly three-fifths of the way out to the edge (Fig. 1). The presence of a moderately well-defined central band of absorbing material across this nucleus closely resembles that which is seen on photographs of external spiral galaxies viewed edge-on The elliptical galaxies, probably the most prevalent type of external system, rarely show any absorbing features at all. *See* GALAXY, EXTERNAL.

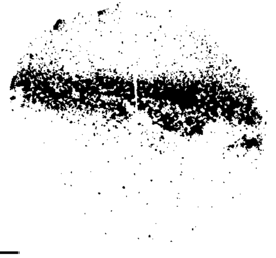


Fig. 1. The Milky Way from Sagittarius to the Southern Cross. This photograph, made in infrared light with a wide-angle camera, shows the region of the Milky Way in the direction of the nucleus of the local galaxy. The dark areas along the Milky Way are clouds of obscuring matter. *(Photograph by A. D. Code)*

Since 1951, when W. W. Morgan first announced that he had succeeded in locating two and possibly three spiral arms by carefully measuring the distances to a number of high-temperature supergiant stars, many of the details of the spiral structure of our galaxy have become apparent. The procedure depends on the observation of objects known to exist only in spiral arms (such as hot supergiant stars) and measurement of their distances with as high a precision as possible. However, because of the strong absorbing effects of interstellar particles, it is not possible to see clearly more than a few thousand parsecs (1 parsec = 3.26 light years = 3×10^{13} km) in the plane of the local galaxy.

Much attention has been given, therefore, to making observations at radio wavelengths where it is possible to penetrate the interstellar medium. In particular, the 21-cm (1420-Mc) emission line of atomic hydrogen, produced by the transition between the two substates of the ground level of hydrogen, has proven to be a most valuable radiation for this work (*see* RADIO ASTRONOMY). Interstellar neutral hydrogen is known to exist in the spiral arms and rarely between them, and 21-cm observations have now given a fairly complete picture of the spiral structure of the galaxy. Astronomers at the Netherlands Foundation for Radio Astronomy, Leiden, Netherlands, and of the Division of Radiophysics, Commonwealth Scientific and Industrial Research Organization, Sydney, Australia, have mapped the spiral arms over nearly the entire Milky Way (Fig. 2).

The constituents of the spiral arms such as hot supergiants, cepheid variable stars, and interstellar material are classified as population I objects. In a more or less spherical haze centered on the galactic nucleus and with a diameter approximately equal to that of the flattened disk containing the spiral arms is a second population comprised of such features as globular clusters, planetary nebulae, and RR Lyrae stars, classified as population II objects. They are also found in the nucleus of the galaxy. *See* STAR.

Fairly reliable distances can be derived for globular clusters and RR Lyrae stars, particularly in the spherical system surrounding the galaxy, because there is no interstellar material present in population II. With an accurate knowledge of the locations of these objects, it is possible to find the distance from the Earth to the center of the nucleus. The best value of this distance is 8300 parsecs (27,000 light years). From radio maps of the galaxy it is found that the diameter of our spiral system is 25,000–30,000 parsecs. The nucleus, although difficult to see because of intervening interstellar material, is probably moderately flattened with the largest diameter measuring perhaps 5000 parsecs. At the distance of the Sun from the center of the galaxy, thickness of the spiral arms is not far from 1000 parsecs. The galaxy and its nucleus are not sharply defined and these figures are all necessarily approximate.

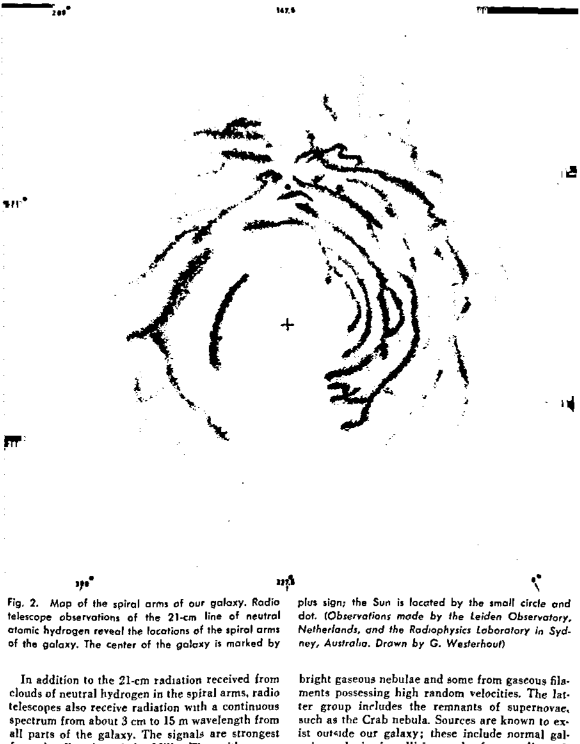


Fig. 2. Map of the spiral arms of our galaxy. Radio telescope observations of the 21-cm line of neutral atomic hydrogen reveal the locations of the spiral arms of the galaxy. The center of the galaxy is marked by plus sign; the Sun is located by the small circle and dot. *(Observations made by the Leiden Observatory, Netherlands, and the Radiophysics Laboratory in Sydney, Australia. Drawn by G. Westerhout)*

In addition to the 21-cm radiation received from clouds of neutral hydrogen in the spiral arms, radio telescopes also receive radiation with a continuous spectrum from about 3 cm to 15 m wavelength from all parts of the galaxy. The signals are strongest from the direction of the Milky Way with a pronounced maximum in the constellation Sagittarius, *the direction of the galactic nucleus. The source of* this radiation is probably largely of the same type observed in synchrotrons, that is, fast-moving electrons in magnetic fields. In addition, nearly 2000 discrete sources of radio waves have been detected, the brighter of which tend to be concentrated along the Milky Way. Most of these sources are yet unidentified. However, some are known to come from bright gaseous nebulae and some from gaseous filaments possessing high random velocities. The latter group includes the remnants of supernovae, such as the Crab nebula. Sources are known to exist outside our galaxy; these include normal galaxies, galaxies in collision, and a few peculiar galaxies. The galaxy has a weak general magnetic *field. which tends to align interstellar crystalline* material causing the light from distant stars to be weakly polarized. The strength of the field in the spiral arms, along which the lines of force run, is estimated by L. Davis and J. L. Greenstein to be 10^{-4} gauss.

Galactic circle. The galactic circle or galactic equator is defined as that great circle on the celes-

tial sphere which best fits the band of the Milky Way. It is inclined at an angle of approximately 62° to the celestial equator, the projection of the Earth's Equator on the celestial sphere. The galactic center lies in the direction of the constellation Sagittarius. It is from this point that galactic longitude is measured eastward in degrees of arc. Galactic latitude is measured north and south from the galactic equator. The north galactic pole is located in the constellation of Coma Berenices.

Galactic rotation and mass. From measurements of the Doppler shifts of absorption lines in the spectra of different regions of spiral galaxies which are seen in orientations other than face-on, it is obvious that these objects are in rapid rotation about an axis perpendicular to their planes and passing through the center of their nuclei. Such a behavior certainly would be expected from the large degree of flattening observed. From the measurements of N. U. Mayall it is known that the central regions of a galaxy rotate as a rigid body; that is, the velocity of rotation increases proportionally to the distance from the center. However, at greater distances from the nucleus, objects move with velocities which decrease as the distance to the center increases, as would be expected if all the mass of the galaxy were concentrated at the center. Our own galaxy must rotate in the same way.

Mayall has also measured the Doppler shifts in the spectra of the twelve galaxies nearest to the Sun and finds that if it is supposed that these galaxies are on the average motionless with respect to the center of our galaxy, the Sun must be revolving about the center of the galaxy with a velocity of 300 km/sec.

By observing the differential motions of objects within our own galaxy, it is possible to arrive at similar results. For example, objects at galactic longitudes 0°, 90°, 180°, and 270° should show on the average no radial velocities with respect to the Earth, because they are either moving at right angles to the line of sight or are moving with the same velocity as the Sun. At a longitude of 45°, however, objects are found to be moving away from the Sun; they are traveling with a higher velocity than the Sun, and their velocity vector is directed away from the Sun. At a longitude of 135°, the situation is reversed because the Sun is moving with the higher velocity, and so on. The radial velocities may be closely represented by the equation

$$\text{Radial velocity} = rA \sin 2l$$

where r is the distance to the object from the Sun, l is its galactic longitude, and A, known as Oort's constant, depends on many factors including the mass of the galaxy and the distance of the Sun

much work on galactic structure and dynamics.

A similar differential effect is expected from measurements of the apparent motions of stars across the line of sight. However, the necessary observations are exceedingly difficult to make.

A thorough study of differential radial velocities leads to several interesting results, one of which is that the velocity of revolution of the Sun about the center of the galaxy is close to 220 km/sec, a value considerably lower than that found by Mayall. However, the distance to the center of the galaxy agrees well with the earlier value given.

The spherical shape of the halo of globular clusters about the galaxy suggests that, on the average, these objects define a system at rest with respect to the spiral system. Indeed the Sun's velocity measured relative to the globular clusters is about 200 km/sec. Perhaps even more significant is the result that the direction of the Sun's motion is almost exactly towards a longitude of 90°, indicating that the Sun is traveling in a nearly circular orbit about the galactic center. The period of time it takes the Sun to make one revolution about the nucleus would then be 2×10^5 years.

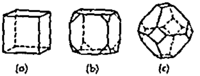
The distance of the Sun from the galactic center and the Sun's period of revolution around the galaxy give the mass of the galaxy, assuming that it all is concentrated at the center. This value is 2×10^{11} times the mass of the Sun, which is 2×10^{15} g. This figure is probably not appreciably different from the true mass of the galaxy, because most of the mass of the galaxy must certainly be in the inner regions. Approximately 90% of this mass is believed to be made up of stars; the remainder comes from the interstellar material (*see* INTERSTELLAR MATTER). In the vicinity of the Sun, however, the interstellar gas and grains perhaps comprise as much as one-half of the mass. Here the density of all matter is slightly less than one-tenth of a solar mass per cubic parsec or 5×10^{-23} g/cm^3.

Comparison with other galaxies. Reliable dimensions of other galaxies will not be available until the distances involved are well determined. On the basis of the best distance measurements, however, it seems that our own galaxy ranks among the largest, brightest, and most massive known. The nearest spiral galaxy to the Sun, the Andromeda nebula, appears to be nearly an identical twin of the Milky Way galaxy. Other neighboring systems, such as the Sculptor galaxy, have total luminosities less than 10^{-5} that of the Milky Way galaxy. In units of the Sun's brightness, the total luminosity of our galaxy is not far from 5×10^9. [W.L.]

Bibliography: B. J. Bok and P. F. Bok, *The Milky Way*, 3d ed., 1957; P. Hanbury Brown and A. C. B. Lovell, *The Exploration of Space by Radio*, 1958; L. Goldberg (ed), *The Structure of the Galaxy*, 1951.

Galena

A mineral having composition PbS. Galena is the principal ore of lead. It crystallizes in the isometric system; cubic crystals frequently truncated by the octahedron, and more rarely by the dodecahedron, are common. The mineral has perfect cubic cleavage, bright metallic luster, and lead-gray color. The hardness is 2.5 (Mohs scale) and the specific gravity is 7.5.



Galena crystals. (a) Cube. (b-c) Octahedral truncations of cube. (From C. S. Hurlbut, Jr., *Dana's Manual of Mineralogy, 16th ed.*, Wiley, 1952)

The theoretical percentages of the elements in galena are lead, 86.6, and sulfur, 13.4. However, analyses usually show the presence of silver, probably as admixtures of silver minerals as argentite, Ag_2S. Because of the silver content, galena in places becomes a valuable source of that element.

Galena is a common mineral found in veins associated with other sulfides, notably sphalerite, pyrite, marcasite, and chalcopyrite, and with gangue minerals such as calcite, quartz, barite, and fluorite. It also occurs in limestones as veins, open-space fillings, or replacement deposits. Galena is widely distributed on all continents of the world and is mined in many countries. Notable localities are Freiberg, Germany; Cornwall and Cumberland, England; Broken Hill, Australia; and in the United States, in southeastern Missouri; the tri-state district of Missouri, Kansas, and Oklahoma; Wallace, Idaho; and various districts in Colorado. *See* LEAD. [C.S.HU.]

Galilean transformations

The family of mathematical transformations used in Newtonian mechanics to relate the space and time variables of uniformly moving (inertial) reference systems.

In the simple case of two similarly oriented cartesian reference frames, moving along their common (x,x') axis, the transformation equations can be put in the form

$$x' = x - vt \qquad y' = y \qquad z' = z \qquad t' = t$$

where x, y, z and x', y', z' are the space coordinates of a given particle and v is the speed of one system relative to the other.

The transformation equations for cartesian reference frames having arbitrary displacements and orientations take the more general form ($x_1 = x$, $x_2 = y$, $x_3 = z$)

$$x'_j = \sum_{k=1}^{3} c_{jk}(x_k - a_k - v_k t)$$
$$t' = t - a_4$$

where a_1, a_2, a_3, a_4, v_1, v_2, v_3 are arbitrary real numbers and the coefficients (c_{jk}) are constants. The matrix $C = [c_{jk}]$ is real and orthogonal, so that it satisfies the condition $C^{-1} = C_t$, where C^{-1} and C_t are the inverse and transposed matrices of C.

The Galilean transformations form a 10-parameter group which can be generated from translations of the space and time coordinates, of the space coordinate frame, and transformations to moving reference frames. *See* FRAME OF REFERENCE. [E.L.HI.]

Gallbladder

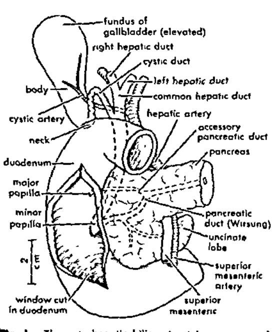
A hollow muscular organ, present in man and most vertebrates, which receives dilute bile from the liver, concentrates it, and discharges it into the duodenum. Although not a vital organ, it is of great importance in man because it stores bile, regulates biliary tract pressures, and, when diseased, causes precipitation of various constituents of the bile as *gallstones*.

Extrahepatic biliary tract. The system of bile ducts lying outside the liver is known as the *extrahepatic biliary tract.*

In man (Fig. 1) right and left hepatic ducts empty into the common hepatic duct which continues to the duodenum as the common bile duct or ductus choledochus. The gallbladder and cystic duct thus appear to be accessory organs and therefore are removable. However, they are converted into main-line structures by the presence of a sphincter at the choledochoduodenal junction. Tonic contraction of this sphincter between meals forces the bile to back up into the gallbladder. *See* LIVER.

In most other vertebrates (Fig. 2) essentially similar relations exist except when the gallbladder is absent, but there is considerable variation in proportion and arrangement of ducts, including the pancreatic ducts. *See* PANCREAS.

Absence of the gallbladder. In certain species, and as a rare anomaly in others, the gallbladder may be absent. The organ first appears in cyclostomes, elasmobranchs, and lungfishes, but not in true bony fishes. It is often absent in birds, but it is always present in amphibians, reptiles, primitive



1. The extrahepatic biliary tract in

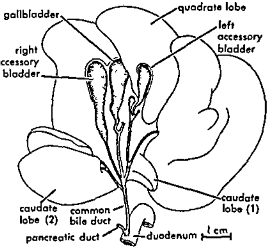


Fig. 2. Accessory gallbladders of ductular type in the cat. (*After E. A. Boyden, Am. J. Anat., 38:177–231, 1926*)

mammals, and their derivative groups, the insectivores, bats, flying lemurs, and primates, and in carnivores. It is absent in many later-originating herbivorous types such as the horse, tapir, rhinoceros, elephant, deer, peccary, and whale. In the giraffe, hippopotamus, and others the gallbladder is disappearing. Conditions are most chaotic in the ocene-developed group of rodents; among these it is missing in kangaroo rats, common rats, porcupines, and some gophers. In man, the most common anomaly is the folded fundus, with an incidence of 2%. Four different types of accessory gallbladders also occur in mammals, tending to be characteristic for a given species, as those of man, the cow, sheep, pig, and cat (Fig. 2).

Development of extrahepatic biliary tract. The way in which this system develops in man is responsible for certain untoward clinical episodes. Both liver and ventral pancreas arise as outgrowths of a diverticulum of the embryonic gut (Fig. 3). Because the ventral pancreas is carried out on the bile duct, the latter becomes embedded in the head of the pancreas. Cancer of this gland can therefore obstruct the flow of bile, causing it to back up into the capillaries of the liver. Another consequence is the creation of a terminal chamber, the ampulla of Vater, into which bile and pancreatic ducts empty. Stones acting as a ball valve may lodge there; or spasms of the sphincter ampullae may result in reflux of bile into the pancreatic duct, a factor in pancreatitis. The left lobe of the ventral pancreas may encircle the gut annular pancreas, with consequent obstruction of the duodenum. In cats such aberrant lobes may grow up alongside the gallbladder, giving rise to a pancreatic bladder. Because of the oblique passage of the ducts through the intestinal wall, herniations or duodenal diverticula of the mucous membrane may develop. *See* LIVER DISORDERS; PANCREAS DISORDERS.

Choledochoduodenal junction. The oblique pathway through the duodenal wall traversed by the common bile duct and pancreatic duct is the choledochoduodenal junction. In man this is surrounded by a musculus proprius which is subject to the same hormonal control as the gallbladder. This sphincter of Oddi, developing from embryonic connective tissue at a later period than the intestinal muscle, differentiates into a sphincter ductus choledochus, a sphincter ductus pancreaticus, and a sphincter ampullae, if an ampulla persists. At the point where the ducts enter the duodenum this sheath receives auxiliary fibers from the intestinal muscle which tie the ducts to the intestine and assist in the erection of the papilla (Fig. 4). Because of this connection some authors have denied the independence of the sphincter. It can act independently, however, and its spasms may result in one type of biliary colic. After removal of the gallbladder a postcholecystectomy syndrome may develop which often can be attributed to an irritable sphincter. Physiologically its most significant component is the sphincter ductus choledochus. Tonic

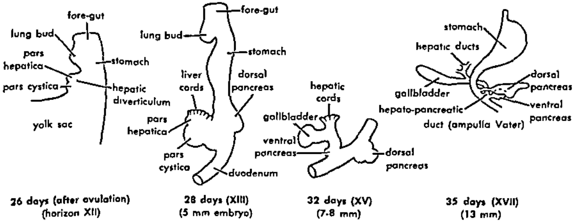


Fig. 3. Development of gallbladder (pars cystica) in human embryos. Roman numerals indicate developmental horizons of Streeter.

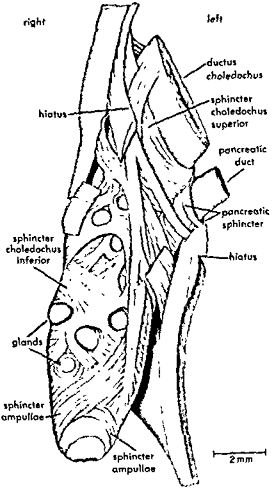


Fig. 4. Human sphincter of Oddi. Upper and lower margins of window in muscle layer of duodenum, through which bile and pancreatic ducts enter major duodenal papilla (Fig. 1). *(After E. A. Boyden, Surg. Gynecol. Obstet., 104(6):641–652, 1957)*

contraction of this causes filling of the gallbladder.

In other mammals the anatomy of the junction has been worked out in only a few species, namely the cat, dog, chimpanzee, and monkey. The first two differ from primates in that the bile duct passes through a funnel in the intestinal muscle before it enters the subterminal area where the special sphincters lie (Fig. 5). Therefore, bile flow in such species is regulated partly by control of intestinal activity. Experimental findings cannot be transferred from such species to man without reservations.

Microscopic structure. The wall of the gallbladder consists of four layers: an epithelial tunica mucosa which lines the hollow viscus, a muscularis which contracts to expel the bile, a vascular perimuscularis, and, where the organ is not attached to the liver, a tunica serosa or peritoneum. The mucous membrane is the layer which concentrates the bile seven to ten times. If the membrane is diseased, this function is lost and the situation may be diagnosed by the Graham-Cole test which depends on the concentration by the gallbladder of radiopaque substances administered orally or intravenously to patients. The surface area of this membrane is increased grossly by rugose folds and microscopically by countless minute cellular processes, the microvilli, each with subsidiary lacelike filaments (Fig. 6). These are the structures that remove water and inorganic salts from the bile. These substances are then passed on into a great plexus of veins which fill the rugae and pass out through the perimuscular tunic.

Comparative physiology. In man, evacuation of the gallbladder is accomplished by a trigger mechanism which is set off by the presence of fatty foods, meat, and hydragogue cathartics in the duodenum and upper jejunum. Absorption of these substances by the mucous membrane results in the formation of cholecystokinin, a hormone which rapidly circulates in the blood stream and simultaneously produces contraction of the gallbladder and relaxation of the sphincter of Oddi. The most effective agent is egg yolk. By calculating the changing volumes of the gallbladder in a series of x-rays, it has been shown that the first phase of contraction is the most important, because two-thirds to three-fourths of the contents are expelled within the first 40 min (Fig. 7). Resorption by the intestine stimulates secretion of bile for hours

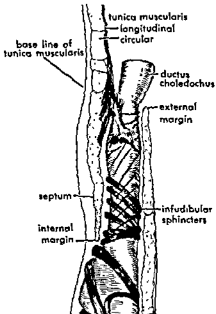


Fig. 5. Choledochoduodenal junction in newborn puppy. (After E. P. Eichhorn and E. A. Boyden, Am. J. Anat., 97:431–459, 1955)

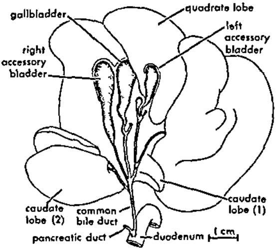


Fig. 2. Accessory gallbladders of ductular type in the cat. (*After E. A. Boyden, Am. J. Anat., 38:177–231, 1926*)

mammals, and their derivative groups, the insectivores, bats, flying lemurs, and primates, and in carnivores. It is absent in many later-originating herbivorous types such as the horse, tapir, rhinoceros, elephant, deer, peccary, and whale. In the giraffe, hippopotamus, and others the gallbladder is . pearing. Conditions are most chaotic in the Eocene-developed group of rodents; among these it is missing in kangaroo rats, common rats, porcupines, and some gophers. In man, the most common anomaly is the folded fundus, with an incidence of 2%. Four different types of accessory gallbladders also occur in mammals, tending to be characteristic for a given species, as those of man, the cow, sheep, pig, and cat (Fig. 2).

Development of extrahepatic biliary tract. The way in which this system develops in man is responsible for certain untoward clinical episodes. Both liver and ventral pancreas arise as outgrowths of a diverticulum of the embryonic gut (Fig. 3). Because the ventral pancreas is carried out on the bile duct, the latter becomes embedded in the head of the pancreas. *Cancer of this gland can therefore* obstruct the flow of bile, causing it to back up into the capillaries of the liver. Another consequence is the creation of a terminal chamber, the ampulla of Vater, into which bile and pancreatic ducts empty. Stones acting as a ball valve may lodge there; or spasms of the sphincter ampullae may result in reflux of bile into the pancreatic duct, a factor in pancreatitis. The left lobe of the ventral pancreas may encircle the gut annular pancreas, with consequent obstruction of the duodenum. In cats such aberrant lobes may grow up alongside the gallbladder, giving rise to a pancreatic bladder. Because of the oblique passage of the ducts through the intestinal wall, herniations or duodenal diverticula of the mucous membrane may develop. *See* LIVER DISORDERS; PANCREAS DISORDERS.

Choledochoduodenal junction. The oblique pathway through the duodenal wall traversed by the common bile duct and pancreatic duct is the choledochoduodenal junction. In man this is surrounded by a musculus proprius which is subject to the same hormonal control as the gallbladder. This sphincter of Oddi, developing from embryonic connective tissue at a later period than the intestinal muscle, differentiates into a sphincter ductus choledochus, a sphincter ductus pancreaticus, and a sphincter ampullae, if an ampulla persists. At the point where the ducts enter the duodenum this sheath receives auxiliary fibers from the intestinal muscle which tie the ducts to the intestine and assist in the erection of the papilla (Fig. 4). Because of this connection some authors have denied the independence of the sphincter. It can act independently, however, and its spasms may result in one type of biliary colic. After removal of the gallbladder a postcholecystectomy syndrome may develop which often can be attributed to an irritable sphincter. Physiologically its most significant component is the sphincter ductus choledochus. Tonic

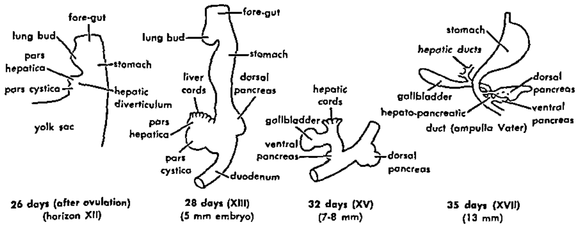


Fig. 3. Development of gallbladder (pars cystica) in human embryos. Roman numerals indicate developmental horizons of Streeter.

concentrate well. The gallbladders of the sheep, goat, cow, and pig appear to manifest little, if any, power of concentrating the bile. Correspondingly, the sphincter offers little or no resistance to the flow of bile. In man and the dog, the gallbladder has both a storage and pressure-regulatory function; in the rabbit, sheep, cow, goat, and pig, only a pressure-regulatory function; and in still others such as the guinea pig, it has no apparent function because its removal results in no grossly demonstrable changes. [E.A.B.]

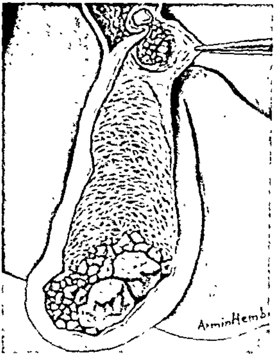
Bibliography: E. A. Boyden, An analysis of the reaction of the human gall bladder to food, *Anat. Record*, 40:147–189, 1928; F. Gorham and A. C. Ivy, Evolutionary contributions to the general function of the gall bladder, *Am. J. Digest. Diseases Nutrition*, 4:792–796, 1938; C. R. Schmidt and A. C. Ivy, The general function of the gall bladder. Do species lacking a gall bladder possess its functional equivalent? *J. Cellular Comp. Physiol.*, 10:365–383, 1937.

Gallbladder disorders

Although not essential to life or health, the gallbladder is the site and source of appreciable suffering and disease in man. With its cystic duct, the gallbladder constitutes a blind-ended, lateral extension of the common bile duct. It is roughly the size and shape of a small pear, and is adherent to the outer surface of the liver. Besides acting as a reservoir for bile, the gallbladder concentrates and otherwise alters its composition. *See* GALLBLADDER.

Gallstones. Gallstones are round, oval, or faceted concretions formed within the gallbladder from the salts and pigment of bile. Such stones may also be formed in any of the bile ducts within or outside of the liver, but the incidence there is low compared to the number originating in the gallbladder. The occurrence of gallstones increases with age, and 20–30% of all elderly adults have one or more such stones. They are particularly common in women, Negroes, and individuals with diabetes. Usually about the size of a pea or a marble, gallstones may be extremely tiny, or so large that a single stone completely fills the gallbladder. They may be composed solely of calcium, cholesterol, or bilirubin, but usually represent combinations or mixtures of these three ingredients precipitated from supersaturated bile. Although the mechanism and reason for their formation is not clearly understood, the major predisposing factors are stasis (prolonged retention of bile in the gallbladder), abnormal composition of the bile (excessive amounts of cholesterol, bilirubin, or calcium), and infection. Passage of a gallstone through the ducts into the duodenum usually produces severe pain, called biliary colic. Lodged in the distal portion of the common bile duct, a stone may obstruct that structure and produce jaundice and eventually cirrhosis of the liver. The most frequent complication of gallstones is cholecystitis. The frequent association of gallstones with cancer of the gallbl[illegible] indicates that they are important in the development of these tumors. Gallstones are rare in animals, although they have been found in nearly all species, and especially cattle.

Cholecystitis. Cholecystitis, or inflammation of the gallbladder, is a common disease in man; as many as 70% of all elderly patients show anatomical evidence of having had one or more attacks. It is nearly always associated with gallstones and is particularly common in obese middle-aged women. It is rare in animals. Most cases are thought to be the result of chemical irritation caused by excessively concentrated bile, which is in turn the result of partial or complete obstruction to the outflow of bile. The gallbladder which is irritated and inflamed in this fashion is susceptible to secondary bacterial infection, and this, in turn, may spread throughout the bile ducts into the liver tissue. The inflamed gallbladder is usually enlarged because of edema of its wall and the presence of pus in its lumen. Perforation may occur, leading to the development of peritonitis or to abnormal communications with the intestine. Prolonged or recurrent episodes of inflammation result in chronic cholecystitis, characterized by thickening and scarring of the wall, contraction, and impairment of normal function. The latter may be detected clinically by administering a radiopaque dye which in a normal gallbladder is concentrated sufficiently to produce a shadow on the x-ray film. The absence of such a shadow indicates a failure of this concentrating ability, and suggests past or present gallbladder disease.



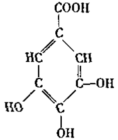
. with numerous faceted gallstones.

Tumors. Malignant tumors are less common in the gallbladder than in several other parts of the intestinal tract, such as the stomach, colon, and pancreas, but comprise as much as 3% of all malignant tumors in man. They are almost invariably associated with the presence of gallstones. Because they produce little in the way of symptoms, and because they very soon invade the liver, these tumors are rarely curable by surgical therapy at the time they are discovered. Similar carcinomas arise in the common bile duct, and in the ampulla of Vater, but there the outlook is somewhat better. In the large bile ducts, tumors usually cause biliary obstruction and jaundice early in their course and may be surgically resectable if the formidable technical problem of reestablishing biliary and pancreatic drainage can be surmounted. Benign tumors of the gallbladder and ducts are rare in man, and in animals both benign and malignant tumors of these sites are extremely uncommon. *See* NEOPLASIA. [M.R.H.]

Bibliography: S. L. Robbins, *Textbook of Pathology*, 1957; H. A. Smith and T. C. Jones, *Veterinary Pathology*, 1957.

Gallic acid

An organic acid with the formula $C_6H_2(OH)_3$-COOH. It is produced commercially by fermentation (*see* FERMENTATION). Gallic acid is used in the engraving, tanning, printing, and pharmaceutical industries. The compound melts and decomposes at 235°C and loses 1 molecule of water at 100°C It has the following structural formula



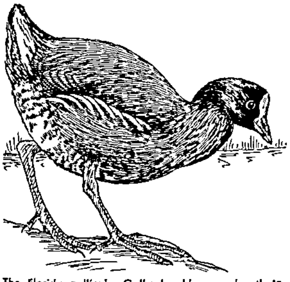
The gallic acid fermentation is the action of filamentous fungi on plant tannins to release gallic acid (3,4,5-trihydroxybenzoic acid). Gallic acid occurs in ester, or glycosidic combination, in gall nuts, in the bark of hemlock and oak, and in the wood of chestnut and quebracho trees. Aqueous tannin-containing extracts of the plant materials are inoculated with specially selected strains of *Aspergillus niger*. The extract usually contains adequate minerals and is bactericidal for most microorganisms so that asepsis is not vital. The inoculated liquor is aerated and agitated under submerged conditions. The mold develops at the expense of utilizable nutrients in the extract, producing an extracellular enzyme, tannase, which hydrolytically liberates gallic acid. The product is crystallized from the filtered fermentation liquor. *See* INDUSTRIAL MICROBIOLOGY. [J.W.F.]

Galliformes

The order of birds of perhaps greatest economic importance to man, including as it does the most important domestic and game birds. Seven families of living Galliformes usually recognized are Megapodiidae (mound birds), Cracidae (guans, curassows), Tetraonidae (grouse), Phasianidae (pheasants, quail), Numididae (guineafowl), Meleagrididae (turkeys), and Opisthocomidae (hoatzins). The one living species of hoatzin is a South American bird of uncertain affinities, notable for the presence of functional wing claws, used in climbing among branches, in the young. Among the pheasants are two species which have been carried by man almost throughout the world. The domestic chicken is a descendant of the red jungle fowl (*Gallus gallus*) of southeast Asia, and has been kept for centuries by both primitive and civilized man. The ring-necked pheasant (*Phasianus colchicus*), a widely distributed Asiatic species, has been introduced in many countries as a game bird. Young galliform birds develop rapidly, and can soon fly. In the mound birds (Megapodiidae) the chick develops to this point while still within the egg and can fly immediately after hatching. Among the grouse, three species of ptarmigan (*Lagopus*) are notable for having evolved a supplementary white winter plumage without counterpart in other birds. *See* AVES. [K.C.P.]

Gallinule

Any of three species of the genus *Porphyula*, found in the warm parts of Africa and America. Species in the United States are the purple gallinule, *P. martinica*, and the Florida gallinule, *Gallinula chloropus*. The purple gallinule is similar to the coot, but is readily distinguished by the yellow-tipped, red bill. The Florida race has a narrow

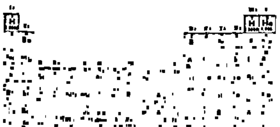


The Florida gallinule, *Gallinula chloropus*; length to 14¾ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

white flank stripe, and the forehead is red, rather than light blue as in the typical race. Gallinules live in the fresh-water swamps of the South, and southward into tropical America. They sometimes wander northward, and have been reported in England and other places far from their natural range. *See* COOT; GRUIFORMES. [J.D.B.]

Gallium

Element number 31, gallium, Ga, was discovered by Lecoq de Boisbaudran in France in 1875. The discovery followed an extended spectroscopic search, instigated because Boisbaudran had noted that a gap occurred in the regularity of the spectral lines between aluminum and indium in the third group of the periodic arrangement of the elements. After examining hundreds of mineral samples, Boisbaudran found the missing spectral lines in a sample of zinc blende from the Pyrenees. In the same year, he isolated a few grams of the metal and named it gallium in honor of France.



Uses. The uses for gallium have been largely experimental. Its great temperature range in the liquid state has prompted its use to some extent in high-temperature thermometers. As an alloy with silver and tin, it is a suitable replacement for amalgam in dental fillings. It has received considerable attention as a heat-exchange medium for nuclear reactors. However, its highly corrosive effect on most metals at elevated temperatures coupled with its high price have discouraged this use.

In the form of the arsenide and the antimonide, gallium has given excellent results as a semiconductor for use in rectifiers and transistors. The excellent semiconductive properties of these compounds, particularly at higher temperatures than are usable with germanium and silicon, give promise of a large future market for high-purity gallium. For this application, gallium is being produced with a purity of 99.9999%.

Because gallium production involves many chemical operations on large quantities of raw material, its price is relatively high ($2–3 per gram). Its production rate has not exceeded a few hundred pounds per year.

Occurrence and extraction. Gallium occurs in nature in only very low concentrations, usually less than 0.01%. It is widely distributed, its abundance having been estimated at 15 g/ton in the earth's crust. This is about equal to the abundance of lead and 30 times that of mercury, but there are no ores sufficiently rich in gallium to make its direct extraction from an ore economically feasible. Germanite, the richest ore, contains an average of only 0.6% gallium. Consequently, gallium is obtained as a by-product from the production of other metals or from waste products.

Following Boisbaudran's discovery, 40 years elapsed before further interest in the recovery of gallium occurred. In 1915 F. G. McCutcheon, in the United States, observed that the lead-containing residues from the refining of zinc smelted from the Tri-State ores would exude drops of a liquid metal following exposure to rain. An analysis of these drops showed that the liquid metal was an alloy containing about 94% gallium and 6% indium. He concluded that sodium and calcium in the residue reacted with the rain water and released the rare metals. Working with these residues, McCutcheon produced the world's first pound of gallium and pioneered in the development of a commercial process for producing gallium from zinc smelter residues. Following chemical concentration of the gallium, the metal is recovered electrolytically from a strongly alkaline solution and finally purified by recrystallization.

In addition to the recovery from zinc smelter residues, gallium is also produced in a state of high purity from the sodium aluminate liquor obtained in the refining of bauxite by the aluminum industry.

Bauxite, the primary source of aluminum, contains gallium in the amount of a few thousandths of one per cent. In the regular Bayer process of refining the bauxite prior to smelting to aluminum, the bauxite is dissolved in caustic soda and pure aluminum oxide trihydrate is precipitated by seeding and cooling the aluminate liquor. After repeated dissolution of bauxite and precipitation of alumina, gallium builds up in the recycled liquor to an equilibrium value of about 0.1 g/liter. This liquor is therefore a relatively rich source of gallium.

Gallium is separated from alumina in the liquor by utilizing the fact that gallium is more acid than alumina. If the aluminate liquor taken at the end of the alumina trihydrate precipitation cycle is treated with lime, most of its alumina can be precipitated whereas the gallium remains in solution. A carbon dioxide treatment of the liquor will then precipitate the gallium along with the fraction of alumina remaining. An alternate procedure (believed to be practiced in Europe) is to precipitate the bulk of the alumina by a slow gassing with carbon dioxide and then to precipitate the gallium and the remaining alumina by a second gassing with carbon dioxide. In either case the final precipitate, a gallia-rich alumina, is dissolved in caustic soda and, following a sodium sulfide treatment to remove heavy metals, is electro[illegible] to obtain metallic gallium.

In Great Britain, gallium is [illegible] from [illegible] ashes. The ash (flue dust) is [illegible]h cau

soda, the cake is dissolved in water, and most of the alumina and silica are removed by controlled precipitation with hydrochloric acid.

Immediately prior to World War II, most of the world's gallium was produced in Germany. Intermediate products from the processing of the copper-bearing schists of the Mansfeld district were used.

In all processes, the ultimate step is electrolysis of either an alkaline or an acid solution of the gallium. In most cases, sodium gallate is the solution electrolyzed. Alumina is present in this solution in substantial amount, but very little enters the gallium. Gallium is deposited at the cathode, which may be platinum, nickel, or the liquid gallium itself. The anode may be carbon, nickel, or platinum. The current density at the cathode is relatively high and the potential is about 5 volts.

France is currently the largest producer of gallium in Western Europe. The French production and that of the USSR are believed to be obtained either from sodium aluminate liquor or from other intermediate products in the refining of bauxite for the aluminum industry.

Physical properties. Solid gallium has a bluish-gray color; the crystals are orthorhombic. Liquid gallium, which resembles mercury, is silver-white with a bright mirror surface. Although the true freezing point of gallium is 29.8°C, the pure liquid has a marked tendency to supercool and may be held in an ice bath for days without crystallizing. However, addition of a single crystal of the solid metal or even ice will cause rapid crystallization. The hardness of the crystal is 1.5 (Mohs scale). The crystals are highly anisotropic, the electrical conductivity, heat conductivity, and coefficient of expansion varying greatly from one crystal axis to another. Liquid gallium wets magnesia but does not wet beryllia or alumina.

The boiling point of gallium is near 2000°C. Tin is the only other metal that is a liquid over such a broad temperature range, and only two metals have melting points lower than that of gallium, namely, mercury (−39°C) and cesium (28.5°C).

Gallium is different from most metals (but similar to water) in that it expands on solidifying. This property, coupled with its freezing point of 29.8°C, makes it necessary to use elastic containers, such as rubber bulbs or flexible plastic bottles for shipment, since alternate freezing and melting of the gallium would cause breakage of a rigid container and loss or contamination of the metal.

Chemical properties. Gallium is chemically similar to aluminum. It is amphoteric but slightly more acid than aluminum; for example, a solution of sodium gallate is more stable than a solution of sodium aluminate. This property is used in the commercial preparation of gallium, as previously described. The normal valence of both aluminum and gallium is 3+, and the two metals form corresponding hydroxides, oxides, and salts. However, it appears that gallia forms only a monohydrate, whereas alumina forms a monohydrate and a trihydrate. Both aluminum and gallium form alums and a long list of organometallic compounds.

The salts of gallium are colorless; they are prepared directly from the metal, since purification of the metal is simpler than purification of the salts. Gallium trichloride is soluble in many organic liquids, including ether, benzene, carbon tetrachloride, and carbon disulfide. Ether extraction from a water solution is used commercially to purify gallium chloride, and its solubility in many organic liquids is the basis for its use as a catalyst in organic synthesis.

Gallium metal can be dissolved by heating with either the caustic alkalies or the common mineral acids. However, the ease of attack decreases with increasing purity of the metal. Most metal impurities can be removed from gallium metal by alternate treatment at ordinary temperatures with nitric acid and hydrochloric acid, with intermediate, thorough washing with water. Nitric acid disperses the metal into small droplets, and hydrochloric acid causes the droplets to coalesce.

Gallium forms alloys readily with many metals. With aluminum, it forms a eutectic that has a freezing point of 26.3°C; all aluminum contains a small amount of gallium as a harmless impurity. Gallium forms low-melting alloys (binary and ternary) with tin and indium, and mixes with tin in all proportions. *See* ALUMINUM; INDIUM; THALLIUM.

[J.R.W.]

Gallstones

Inorganic nodules formed in the gallbladder or biliary tubes. They may be composed of calcium, cholesterol, or bilirubin, or a combination of these.

Pure stones account for 10% of all gallstones and are composed of cholesterol, calcium bilirubinate, or calcium carbonate, in that order. The typical cholesterol stone is usually single, round to oval in shape, white to yellow in color, and has a smooth or finely granular surface. The size may be upward of 1 in. in diameter. *See* CHOLESTEROL.

Mixed types comprise 80% of all gallstones and are most often associated with cholecystitis. Two and sometimes all three basic substances are present in the form of multiple, small, faceted stones that are whitish-, grayish-, or brownish-black.

The remaining 10% of gallstones are the combined forms which begin usually as a pure stone upon which other basic materials are deposited.

The exact cause of cholelithiasis, or gallstones, is unknown but contributing factors include infections, abnormal bile production, and poor biliary drainage. Gallstones are far more frequent than is commonly supposed since many are either asymptomatic or are passed into the duodenum without causing a medical or surgical crisis. Over one-fourth of all autopsies reveal the presence of gallstones, according to some studies. Obstruction by stones may induce damage to the liver, pancreas, biliary system, and related structures either directly or through concomitant inflammation.

[E.G.ST.]

Galvanizing

The generic name for any of several techniques for applying thin coatings of zinc to iron or steel objects to protect the ferrous metal from corrosion and to improve its appearance. Hot dipping is widely practiced with mild steel sheet for garbage cans and corrugated iron pipe, and with wire for fencing. The electrolytic process (also called cold galvanizing) is also used for wire as well as for other applications; it saves on zinc but takes more time and makes a less durable coating. Sherardizing is a dry process of coating small objects such as nails under heat in a revolving drum filled with zinc dust. Nails are also frequently galvanized by a process that is a cross between hot dipping and sherardizing, and that makes silvery bright nails.

Molten zinc is sometimes sprayed on large objects in place, but this is not a true galvanizing because the zinc does not alloy with the ferrous metal but acts more as a coat of metal paint. *See* METAL COATINGS. [C.CO.]

Galvanomagnetic effects

Electrical and thermal phenomena occurring when a current-carrying conductor or semiconductor is placed in a magnetic field. The galvanomagnetic effects are closely related to the thermomagnetic effects (*see* THERMOMAGNETIC EFFECTS). Both sets of effects yield important information on the band structure of metals and semiconductors and on the nature of the conductivity process.

Let the electric current density j be transverse to the magnetic field H_z, for example, along x. Then the following transverse-transverse effects are observed:

1. Hall effect, an electric field along y,

$$E_y = Rj_xH_z$$

where R is the Hall coefficient. According to experimental conditions, R may be the adiabatic or the isothermal Hall coefficient. *See* HALL EFFECT.

2. Ettingshausen effect, a temperature gradient along y,

$$(\partial T/\partial y) = Pj_xH_z$$

where P is the Ettingshausen coefficient.

Also the following transverse-longitudinal effects are observed:

3. Transverse magnetoresistance, an electrical potential change along x. *See* MAGNETORESISTANCE.

4. Nernst effect, a temperature gradient along x.

Let the electric current density j be along H. Then, the most important effect is longitudinal magnetoresistance, or an electrical potential change along H.

These effects arise from the Lorentz force on moving charges in a magnetic field. They are complicated by interaction between these charges and the lattice in which they move. For this reason, the signs of R and P may be negative or positive, depending upon composition and in many cases on temperature; the magnitudes show tremendous range. [E.A.; F.K.]

Bibliography: S. Flügge (ed.), *Handbuch der Physik*, vol. 20, 1957; F. Seitz and D. Turnbull (eds.), *Solid State Physics*, vol. 5, 1957.

Galvanometer

A device for indicating very small electric currents. Although the deflection of a galvanometer results from current in the moving coil, the voltage in a closed circuit producing this current is frequently the quantity of interest to the user. Galvanometers may also be used ballistically to integrate a transient current, as from the discharge of a capacitor, or a transient voltage as produced when a coil moves relative to a magnetic field. *See* CURRENT MEASUREMENT; VOLTAGE MEASUREMENT.

While great sensitivity is desirable, reasonably rapid and controlled response is also required so that readings may be taken promptly and accurately. The degree to which the movable system is damped is thus of great importance.

The moving systems of many galvanometers carry a small mirror which deflects a light beam to give an indication. This serves as a means for amplifying small motions, and numerous variations have been devised to obtain optimum results in limited space.

D'Arsonval galvanometer. This is the most common type, and is widely used. Its indicating system consists of a light coil of wire suspended from a copper or gold ribbon a few thousandths of an inch wide and less than 0.001 in. thick. This coil, free to rotate between the poles of a permanent magnet, carries a small mirror which serves as an optical pointer and indicates the coil position by reflecting a light beam onto a fixed scale. This arrangement is shown in Figs. 1 and 2. Current is conducted to and from the coil by the suspension ribbons. The torque which deflects the indicating element is produced by the reaction of the coil current with the magnetic field in which it is suspended.

Sensitivity and transient response. Voltage sensitivity is of importance in applications where the galvanometer serves as the detector of unbal-

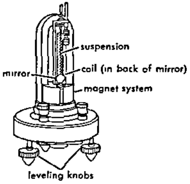


Fig. 1. Light-beam galvanometer. (*From D. M. Considine, ed., Process Instruments and Controls Handbook, McGraw-Hill, 1957*)

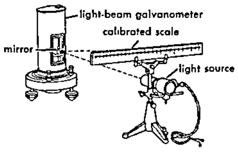


Fig. 2. Light-beam galvanometer showing detached scale and light source for laboratory work. *(From D. M. Considine, ed., Process Instruments and Controls Handbook, McGraw-Hill, 1957)*

ance in a bridge or potentiometer network. Indeed, the response of a galvanometer to voltage is more often of interest than is the current required to produce a specified deflection. With a given magnetic field and coil size, the response to current increases with the number of turns in the coil, so that high current sensitivity is achieved by using a coil having many turns of fine wire, which results in a high resistance. When voltage response is of interest, however, the resistance of the circuit in which the galvanometer operates must be taken into account, and voltage sensitivity is generally greater in a galvanometer of low resistance, that is, with a few turns of relatively coarse wire.

The nature of the response of the galvanometer must always be considered in any application. The energy of motion in the indicating system must be dissipated and the system brought to rest in order for the deflection to be evaluated. This process, common to all indicating systems in which equilibrium must be achieved between a driving torque and a restoring torque, is called damping.

In a galvanometer the driving torque results from the reaction between the magnetic field and the current-carrying conductor, and the restoring force from the stiffness of the suspension as it is twisted by the motion of the deflecting element. The resistance of the circuit in which the galvanometer operates is of major importance in damping. The relative motion between the coil and magnetic field generates an emf that produces a current in the closed circuit of which the coil is a part; energy is dissipated as heat in the resistance of this circuit. The rate of energy dissipation is directly proportional to the square of the generated emf and is inversely proportional to the circuit resistance. Thus if the circuit resistance is too large, the energy of motion is dissipated slowly, the moving system will oscillate about its equilibrium position with gradually diminishing amplitude, and the galvanometer is said to be under-damped. If the circuit resistance is too small and energy is dissipated too rapidly, motion is sluggish and the indicating system creeps toward its equilibrium position. The galvanometer is then said to be over-damped. For a particular intermediate value of circuit resistance, the common limit of the under- and over-damped cases, the galvanometer is critically damped; the value of the resistance in this situation is called the critical damping resistance.

A minimum response time for the deflection of a galvanometer to which an unvarying voltage is suddenly applied will occur when the galvanometer is slightly under-damped, that is, when the circuit resistance is a little greater than the critical damping value. Practically, however, response time is not seriously affected over a range of circuit-resistance value from critical to perhaps 30% greater than critical. Thus a galvanometer selected for a specific application should be chosen with a critical damping resistance that is reasonably well matched to the resistance of the circuit in which it is to be used. Preferably, the circuit resistance should be somewhat greater than that required for critical damping, so that the motion may be slightly under-damped. If the circuit resistance differs greatly from the desired value, it must be adjusted by the addition of series resistance, if too low, or by parallel resistance across the galvanometer terminals, if too high, to obtain proper response characteristics. Even so, such adjustments decrease the operating sensitivity of the galvanometer.

The sensitivity of modern galvanometers ranges up to 1 mm of deflection, on a scale 1 m distant from the mirror, for a current of 0.00001 μa (10^{-11} amp). Such a galvanometer may have a coil resistance of 800 ohms and a critical damping resistance of 100,000 ohms. The voltage response of this instrument amounts to 1 mm/μv at critical damping. A galvanometer by the same maker, designed for voltage sensitivity, has a coil resistance of 20 ohms, a critical damping resistance of 30 ohms, and a response of 1 mm for 0.05 μv in the critically damped circuit. It will be seen from these examples that a large response to current is associated with large coil resistance and high critical damping resistance, whereas voltage response is associated with low coil resistance and low critical damping resistance.

Mechanical oscillation. The indicating system may oscillate because a perfectly stable galvanometer mounting is lacking. The coil of a conventional d'Arsonval galvanometer, designed for fixed mounting on a wall or platform, is generally hung from a metal suspension ribbon which supports the weight of the moving system and provides almost the entire restoring torque. This upper suspension also acts as one of the conductors to introduce current into the suspended coil. The second current lead is a loosely coiled ribbon filament leading from the lower end of the moving coil to the fixed frame. Its contribution to torque is minor. As a result the system behaves like a pendulum. Thus any motion of the upper suspension support will set the galvanometer coil into motion; indeed, the motion of the coil may be many times that of the suspension support. Coil oscillations caused by mechanical disturbances may persist for a consid-

erable time if the motion is linear rather than rotational in the magnetic field, because such a motion does not generate any net emf in the coil and consequently there is no circuit damping to absorb the energy of the motion. Delicate suspension systems of this sort make the galvanometer particularly susceptible to mechanical disturbances, and the instrument should be placed on a special platform or mounting system designed to attenuate mechanical disturbances.

Portable galvanometers. Portable and some wall-type instruments have their coils suspended between upper and lower taut ribbons so that the motion is restricted to rotation about the coil axis. Such taut-suspension systems, however, are stiffer than those having the lower suspension loosely coiled, and such galvanometers are usually less sensitive. Pivoted galvanometers also have their motion confined to rotation. In both of these types the moving system is mechanically balanced around the axis of rotation, and the effect of mechanical disturbances is much reduced.

Ballistic galvanometer. The conventional galvanometer may also be used ballistically to integrate a transient current or voltage pulse of short duration. The electric charge in a capacitor may be measured by allowing it to discharge through a galvanometer, observing the magnitude of the first swing of the moving system from its rest position. A galvanometer's response to current of short duration is called its coulomb sensitivity. If the current pulse is short, the deflection of the moving system is proportional to the time integral of the current, that is, to the electric charge in the pulse. Similarly, a voltage pulse rising from a change in a magnetic field linking a coil may be integrated by a galvanometer. The time integral of voltage induced in the coil is proportional to the ballistic deflection of the galvanometer which is a measure of the total change in coupling between field and coil. This response defines the emf-time (or flux-linkage) sensitivity of the galvanometer.

Sensitivities of four different types thus describe the performance of a galvanometer in its various applications. Current sensitivity is the only one defined solely by the design constants, that is, strength of magnetic field, area turns of the coil, and stiffness of the suspension. Voltage, coulomb, and emf-time sensitivities are all functions of circuit resistance, as well as galvanometer design constants. Response to transient current increases with increasing circuit resistance, whereas both steady-state and transient voltage response increase with decreasing circuit resistance.

Other galvanometers. Although other types are possible, they are not much used in present-day engineering. All types make use of the reaction between a magnetic field and a current-carrying conductor to achieve a measure of the current, but it is not necessary that the conductor be part of the moving element. The reaction is similar and the sensitivity may be equally great if a magnet system is used as the moving element and the current is carried by fixed coils. The chief advantage of this alternative construction is that the suspensions are not required to carry current, and may be extremely fine quartz fibers with reduced stiffness. Thus the reduced driving torque available in the moving-magnet fixed-coil arrangement is compensated by reduced suspension stiffness. The principal disadvantage of such a galvanometer is that its response is more affected by changes in the ambient magnetic field than is the d'Arsonval galvanometer. For reasonable performance a moving-magnet galvanometer must be elaborately shielded from external fields by a series of nested enclosures of high-permeability material, and for best results the construction must also be astatic—that is, the magnet and coil system must be arranged in matched pairs in such a manner that the torques resulting from a uniform external field will oppose and neutralize each other; the torques resulting from the interaction of the current-carrying conductors and suspended magnets are additive. Despite the high sensitivity that can be achieved with this arrangement, it is little used in modern practice because of the difficulty of realizing adequate magnetic shielding.

Indicating means. Several types of indication are available to expand the scale of a galvanometer so that smaller currents or voltages, or smaller changes in their values, may be observed. When a light-beam pointer is used to indicate coil motion, the distance between the galvanometer mirror and scale may be increased. Thus for a given angular motion of the coil, the linear distance traversed by the indicating light spot on the scale is increased. A limit is usually reached in the discrimination of small displacements on the expanded scale as a result of defects in the galvanometer mirror or aberrations in the lens system used to focus the light beam on the scale. Even if the optical system were perfect, however, a limit would still be imposed by diffraction. The galvanometer mirror may be considered as a limiting aperture in the optical system, and its diameter determines the width of the diffraction pattern in the focused image of the light source. The practical limit of resolution is usually reached with a light-beam pointer between 2 and 3 m in length; the theoretical resolution, with a mirror 1 cm in diameter, is reached at about 4 m.

An optical train consisting of two or more elements of appropriate focal length may be used instead of a simple lens to focus an image of the source on the scale. This has the advantage of increasing the angular displacement of an image point in a given physical distance, permitting a more compact arrangement. Actually, such an optical train may be folded, using mirrors of appropriate curvature as the optical elements, so that a 2–3-m effective pointer length can be built into a box no more than 30 cm deep. This arrangement has been used in portable galvanometers. However, any such optical train is subject to the same diffraction limit of resolution, imposed by the mirror [illegible]ing aperture of the system.

Another method of increasing angular displacement of the indicating light spot involves the use of multiple reflections from the galvanometer mirror. For a single reflection from a moving mirror, the angular displacement between the incident and reflected beam is double the angle through which the mirror turns. If the beam is reflected by a stationary mirror back to the moving mirror, the angular displacement of the beam reflected from the moving mirror is again doubled. Thus by repeated reflections between the galvanometer mirror and a fixed mirror, the angular motion of the galvanometer can be greatly magnified. This method has the advantage that the diffraction limit of resolution is a function of total path length so that, if the distance between fixed and moving mirrors is kept small, greater magnification of motion and higher angular resolution can be achieved before reaching the limit imposed by diffraction, than with the use of a multiple-lens system.

The human eye is limited in sensing small motions because it must differentiate between light and dark areas at the edge of a line image on the galvanometer scale. Smaller image displacements can be observed by using devices such as photocells which will integrate the total light received. In practice, the light reflected from the galvanometer mirror may be split into two beams with a biprism or other suitable arrangement, and these beams directed to separate photocells. As the galvanometer coil turns, the light received on one cell increases while the other decreases. If the cells are arranged in a circuit so that the difference in their currents may be observed either directly or through an amplifier, exceedingly small motions of the galvanometer mirror may be sensed. In such an arrangement it is frequently possible to approach rather closely the theoretical limit of resolution of angular motion imposed by the Brownian motion of the coil, or by the Johnson noise of the circuit connected to the galvanometer, whichever limit is first reached. In a low-resistance galvanometer this theoretical resolution may be about 0.001 μv.

Other methods than those of optics are possible for sensing extremely small angular displacements of a galvanometer coil, and hence very small changes in current or voltage. Examples are arrangements in which the coil motion is indicated by changes in inductive or capacitive coupling between the moving element and another element whose position is fixed. Such a system has the advantage, in principle, that the galvanometer may be completely enclosed whereas a window is needed for an optical system; there is, of course, the disadvantage that an auxiliary network is needed for sensing the change in inductance or capacitance. *See* AMMETER; ELECTRICAL MEASUREMENTS; VOLTMETER. [F.K.H.; J.H.MI.]

Bibliography: D. M. Considine (ed.), *Process Instruments and Controls Handbook*, 1957; F. K. Harris, *Electrical Measurements*, 1952; F. A. Laws, *Electrical Measurements*, 2d ed., 1938.

Game theory

The theory of games of strategy can briefly be characterized as the application of mathematical analysis to abstract models of conflict situations. The first such models analyzed by the theory were parlor games such as chess, poker, and bridge. Since then, models arising from the behavioral sciences such as economics, sociology, and political science have been analyzed. Game theory is used in or closely connected to other areas such as linear programming, statistical decisions, management science, operations research, and military planning. In certain areas, the language and concepts of the theory are sometimes used even though the corresponding mathematics is not. *See* LINEAR PROGRAMMING; OPERATIONS RESEARCH.

Games in extensive form. Parlor games are specified by a list of rules such as those contained in Hoyle's well-known book. To be playable, a game must have the following properties: (1) there must be a way of starting the game; (2) there must be a well-defined list of legal (that is, permissible) moves for any possible situation that can be reached in the game; (3) at each move, exactly one of the players must be assigned to make the choice, or else the choice is made by a chance device (such as rolling dice or a spinning pointer); (4) after a finite number of moves, the game is terminated; a winner is declared or payments are exchanged among the players, or both. Any other conflict situation that satisfies mathematical axioms similar to these rules can also be considered a game and the analysis of game theory applied to it.

The players of a game are called persons, and such a person may actually consist of one or more people (for instance, in bridge the pairs of partners, east-west and north-south, each make up a player in the game; it is considered a two-person game). Thus a person in a game may be considered synonymous with a team. Chance moves occur when hands are dealt from a shuffled pack, dice are rolled, or pointers are spun. One says that all chance moves are allotted to the chance player—a fiction that is useful in abstracting properties of games.

When specified by a list of rules a game is said to be in extensive form. For mathematical purposes it is convenient to have games in mormalized form, and for that, the idea of a pure strategy is needed.

A pure strategy for a player (not the chance player) is a complete list of choices of legal moves that he will make for every possible situation that he can find himself in during the game. This is a much more complete list of decisions than that commonly called a strategy. For instance, in tic-tac-toe a small part of a single pure strategy would be, "If he moves to the center, I will move to the upper right-hand corner; but if he moves to the upper right-hand corner I will move to the center . . . ; if he moves first to the center, and I to the upper right-hand corner, and he to the lower right-hand corner, then I move to the upper left-hand corner.

. . ." The number of pure strategies in a game can be astronomical even for childish games such as tic-tac-toe, as any reader who tries to enumerate them will find for himself. Because of the enormous number of pure strategies, the actual applications of game theory even to parlor games have been severely limited by computational difficulties. Simplified versions of the games have been developed for which computations have been completely carried out.

Games in normalized form. After players have chosen pure strategies in a game, they need not physically play the game. Instead they could hand their strategies to a neutral person, or umpire, who could then carry out their instructions and make the moves they would have made. This intuitively obvious idea leads naturally to the normalized form of the game.

Assume for the moment there are no chance moves in the game, that is, that there are n real players but no chance player. Denote by $s_1, s_2, \ldots, s_n$ specific pure strategies for players $1, 2, \ldots, n$, respectively. Given these, the game must be played in exactly one way and a unique outcome will result. Let $P_i(s_1,s_2, \ldots ,s_n)$ be the monetary outcome to player i for this play of the game. (In case only a winner is announced, but no money is exchanged, it can be arbitrarily required that the losers pay one unit each and the winners divide the losers' payments equally, and in this way make a monetary assignment.)

Before the effect of the chance player can be introduced, the important concept of mathematical expectation must be explained. Suppose that $O_1, O_2, \ldots, O_k$ are mutually exclusive monetary outcomes of a chance event and suppose further that they happen with probabilities $p_1, p_2, \ldots, p_k$ where $p_i > 0$ and $p_1 + p_2 + \cdots + p_k = 1$. Then the mathematical expectation E of the chance event is defined to be the sum $E = p_1O_1 + p_2O_2 + \cdots + p_kO_k$.

As an example, suppose that one tosses a coin and is paid \$3 if heads turns up and loses \$1 if tails turns up. If it is assumed that the coin is balanced, each event will happen with probability 1/2 and the expectation is then $3(1/2) - 1(1/2) = 1$. Note that the expectation is not equal to either of the payoffs that can actually be obtained on a single toss of the coin. It can be interpreted as the average amount obtained if the event is repeated many times.

If there are chance moves in the game, a set of pure strategies, one for each player, will not determine a unique outcome of the game but merely a set of possible outcomes. These outcomes will be mutually exclusive and have probabilities depending on the chance moves associated with their occurrence. Hence in this case one can let $P_i(s_1,s_2, \ldots ,s_n)$ be the expected payoff to player i for each $i = 1, 2, \ldots, n$.

Now the normalized form of a game is defined as the list of all expected payoffs to each player for every possible combination of pure strategies. In the case of two-person games it is most convenient to list these in tables called matrices.

Matching pennies. Two players R and C compare pennies, with R trying to match C. Each player can choose either heads or tails. Each has two pure strategies, "choose heads," H, or "choose tails," T. The game can be summarized as follows:

	C chooses	
R chooses	H	T
H	1	-1
T	-1	1

(1)

The entries in the table represent the payments by C to R. Thus, if both R and C choose heads, R matches C and receives one penny from him; but if R chooses heads and C tails, then R receives -1 from C, that is, he pays one penny to C. The other two entries are determined similarly. The solution of this game will be given below.

The Mississippi gambler. A passenger on a Mississippi river boat was approached by a flashily dressed stranger (the gambler) who offered to play the following game: "You take a red ace and a black deuce and I'll take a red deuce and a black trey; we will simultaneously show one of the cards; if the colors match I'll pay you and if the colors don't match you pay me; moreover if you play the red ace we will exchange the difference of the numbers on the cards; but if you play the black deuce we will exchange the sum of the numbers. Since if you lose you will pay me either \$2 or \$4 and if I lose I will pay you either \$1 or \$5, the game is obviously fair."

The proposed game is most easily summarized as follows:

	Gambler chooses	
Passenger chooses	Red 2	Black 3
Red 1	1	-2
Black 2	-4	5

(2)

The payoffs listed are those received by the passenger. It is shown below that this game is not fair, and that the gambler has a substantial advantage.

Simplified poker. The deck used in simplified poker has only two kinds of cards, in equal numbers, high (H) and low (L). For instance, a bridge deck could be used with red cards high and black cards low. There are two players R and C and each player antes an amount a and is dealt one card which is his hand. It is easy to see that there are only four ways to deal the hands in the game, namely, (H,H), (H,L), (L,H), and (L,L), where R gets the first card of the pair and C the second.

After the deal, R has two alternatives, namely to see or to raise by adding an amount b to the pot.

Another method of increasing angular displacement of the indicating light spot involves the use of multiple reflections from the galvanometer mirror. For a single reflection from a moving mirror, the angular displacement between the incident and reflected beam is double the angle through which the mirror turns. If the beam is reflected by a stationary mirror back to the moving mirror, the angular displacement of the beam reflected from the moving mirror is again doubled. Thus by repeated reflections between the galvanometer mirror and a fixed mirror, the angular motion of the galvanometer can be greatly magnified. This method has the advantage that the diffraction limit of resolution is a function of total path length so that, if the distance between fixed and moving mirrors is kept small, greater magnification of motion and higher angular resolution can be achieved before reaching the limit imposed by diffraction, than with the use of a multiple-lens system.

The human eye is limited in sensing small motions because it must differentiate between light and dark areas at the edge of a line image on the galvanometer scale. Smaller image displacements can be observed by using devices such as photocells which will integrate the total light received. In practice, the light reflected from the galvanometer mirror may be split into two beams with a biprism or other suitable arrangement, and these beams directed to separate photocells. As the galvanometer coil turns, the light received on one cell increases while the other decreases. If the cells are arranged in a circuit so that the difference in their currents may be observed either directly or through an amplifier, exceedingly small motions of the galvanometer mirror may be sensed. In such an arrangement it is frequently possible to approach rather closely the theoretical limit of resolution of angular motion imposed by the Brownian motion of the coil, or by the Johnson noise of the circuit connected to the galvanometer, whichever limit is first reached. In a low-resistance galvanometer this theoretical resolution may be about 0.001 μv.

Other methods than those of optics are possible for sensing extremely small angular displacements of a galvanometer coil, and hence very small changes in current or voltage. Examples are arrangements in which the coil motion is indicated by changes in inductive or capacitive coupling between the moving element and another element whose position is fixed. Such a system has the advantage, in principle, that the galvanometer may be completely enclosed whereas a window is needed for an optical system; there is, of course, the disadvantage that an auxiliary network is needed for sensing the change in inductance or capacitance. *See* AMMETER; ELECTRICAL MEASUREMENTS; VOLTMETER. [F.K.H.; J.H.MI.]

Bibliography: D. M. Considine (ed.), *Process Instruments and Controls Handbook*, 1957; F. K. Harris, *Electrical Measurements*, 1952; F. A. Laws, *Electrical Measurements*, 2d ed., 1938.

Game theory

The theory of games of strategy can briefly be characterized as the application of mathematical analysis to abstract models of conflict situations. The first such models analyzed by the theory were parlor games such as chess, poker, and bridge. Since then, models arising from the behavioral sciences such as economics, sociology, and political science have been analyzed. Game theory is used in or closely connected to other areas such as linear programming, statistical decisions, management science, operations research, and military planning. In certain areas, the language and concepts of the theory are sometimes used even though the corresponding mathematics is not. *See* LINEAR PROGRAMMING; OPERATIONS RESEARCH.

Games in extensive form. Parlor games are specified by a list of rules such as those contained in Hoyle's well-known book. To be playable, a game must have the following properties: (1) there must be a way of starting the game; (2) there must be a well-defined list of legal (that is, permissible) moves for any possible situation that can be reached in the game; (3) at each move, exactly one of the players must be assigned to make the choice, or else the choice is made by a chance device (such as rolling dice or a spinning pointer); (4) after a finite number of moves, the game is terminated; a winner is declared or payments are exchanged among the players, or both. Any other conflict situation that satisfies mathematical axioms similar to these rules can also be considered a game and the analysis of game theory applied to it.

The players of a game are called persons, and such a person may actually consist of one or more people (for instance, in bridge the pairs of partners, east-west and north-south, each make up a player in the game; it is considered a two-person game). Thus a person in a game may be considered synonymous with a team. Chance moves occur when hands are dealt from a shuffled pack, dice are rolled, or pointers are spun. One says that all chance moves are allotted to the chance player—a fiction that is useful in abstracting properties of games.

When specified by a list of rules a game is said to be in extensive form. For mathematical purposes it is convenient to have games in mormalized form, and for that, the idea of a pure strategy is needed.

A pure strategy for a player (not the chance player) is a complete list of choices of legal moves that he will make for every possible situation that he can find himself in during the game. This is a much more complete list of decisions than that commonly called a strategy. For instance, in tic-tac-toe a small part of a single pure strategy would be, "If he moves to the center, I will move to the upper right-hand corner; but if he moves to the upper right-hand corner I will move to the center . . . ; if he moves first to the center, and I to the upper right-hand corner, and he to the lower right-hand corner, then I move to the upper left-hand corner.

Matrix games. A matrix is a rectangular array of numbers such as those of tabulations (1) to (4). The matrices in (1) and (2) are called 2-by-2 matrices because they have 2 rows and 2 columns. The matrix in (3) is 4-by-4. In general, a matrix is called m-by-n, if m is the number of its rows and n the number of its columns. Any two-person zero-sum game can be written in normalized form as a matrix game. The matching pennies and simplified poker games discussed above are examples.

Conversely, any matrix with numerical entries can be considered a two-person game between player R who chooses a row of the matrix and player C who chooses a column. After making their choices, C pays R the entry in the common row and column they have chosen (in case the entry is negative, R must pay C the amount).

A matrix then provides an abstract mathematical model for the study of two-person zero-sum games. What is meant by the solution to such a game? Consider the matrix of (5) Player R (the row player) will receive the positive amounts of the matrix and must pay out the negative ones. He is called the maximizing player. Similarly, C (the column player) must pay out the positive amounts and will receive the negative ones, so he is called the minimizing player. In (5), player R would like to get the 5 entry, but to do so would have to choose the second row. And if he did this, player C would

	C chooses	
R chooses	1	2
	−3	5

(5)

choose the first column and R would lose 3. Hence, R plays the first row and C, who wants to minimize his payments, will still choose the first column. The result is that R will receive one dollar from C.

The example of (5) is that of a strictly determined game, that is, a game in which there is a matrix entry that is simultaneously the minimum entry in its row and the maximum entry in its column. Such an entry is called a saddle-point entry of the matrix. Any matrix game that has a saddle-point entry has a solution, namely, that the players should choose rows or columns that contain such saddle-point entries. The value v of such a game is the saddle-point entry itself. In a strictly determined game the row player can assure himself of at least v by playing an optimal strategy. To establish this, note that if he chooses a row whose minimum entry is maximum, then the least payoff he can get is v, the minimum entry of this row. By a similar argument, the column player can assure himself of not losing more than v by playing an optimal strategy.

However, the matrix games of (1) and (2) do not have saddle-point entries as the reader can quickly check. How can such games be solved? Perhaps the matching pennies example of (1) suggests an answer. It is quite common for players of this game to randomize their choice of heads or tails by flipping the coin in the air before choosing. In this way they do not know themselves exactly what choice they will make but leave it to a chance device. Hence, their opponent cannot discover in advance their choice and cannot take advantage of such knowledge. Extensions of this idea give a solution concept to any game, strictly determined or not.

By a mixed strategy for a player in a matrix game is meant a set of probability weights, that is nonnegative numbers that add up to 1. There are exactly as many probability weights as the player has pure strategies. Thus, in an m-by-n matrix game, a mixed strategy for player R is a set of numbers $p_1, p_2, \ldots, p_m$ with $p_i \geqq 0$ and $p_1 + p_2 + \cdots + p_m = 1$. Similarly, the set of numbers $q_1, q_2, \ldots, q_n$ with $q_j \geqq 0$ and $q_1 + q_2 + \cdots + q_n = 1$ is a mixed strategy for player C. In using a mixed strategy, the player operates a chance device that chooses among the alternatives with probabilities equal to the corresponding weights. For instance, the device of flipping a coin in a matching pennies game has already been mentioned. In other games, the random choice of a pure strategy may be made by rolling dice, spinning pointers, or choosing random numbers.

A mixed strategy has the desirable property that a player's opponent cannot find out exactly what his choice will be, because he does not know himself until he has operated the chance device. But even more is true. The use of a mixed strategy by a player introduces a chance move into the game and eliminates all his personal moves. Hence, its effect in the game can be accounted for by computing each player's expected value, given a choice of mixed strategy for each. Let p stand for player R's mixed strategy and q for player C's. Let $E(p,q)$ be the expected payoff to R for these choices; then $-E(p,q)$ is C's expected payoff. Therefore if only the payoff $E(p,q)$ is considered, it is evident that R wants to choose p so as to maximize this amount and C wants to choose q to minimize it.

To analyze the game further, consider two modified games, one in which player C makes his choice of mixed strategy after R has announced his mixed strategy and the other in which R makes his choice after C's announcement. If C chooses after R, he will select a q that minimizes the result; because R knows that C will do this, he selects his p initially to maximize the eventual result, that is, in this case p and q are chosen so that

$$\max_p \min_q E(p,q)$$

is obtained. In the other game, in which R chooses after C, p and q will be chosen so that

$$\min_q \max_p E(p,q)$$

is obtained. It is not hard to show that

$$\text{in } E(p,q) \leq \min_q \max_p E(p,q)$$

for any matrix game. It is obvious that if R chooses his strategy according to the max-min rule, he is playing very conservatively and is protecting himself from the worst that C can do. Similarly for C and the min-max strategy.

The celebrated min-max theorem of John von Neumann, proved in 1928, states that even more than (6) is true; namely, that for any matrix game there exist mixed strategy vectors $\mathbf{p}^0$ and $\mathbf{q}^0$ so that equality in (6) holds. In other words,

$$E(\mathbf{p}^0,\mathbf{q}^0) = \max_p \min_q E(p,q) = \min_q \max_p E(p,q) = v \quad (7)$$

It is this theorem that makes the study of games a mathematical theory rather than just a descriptive, intuitive theory. The strategies $\mathbf{p}^0$ and $\mathbf{q}^0$ are called optimal mixed strategies for the players R and C, respectively, and $v = E(\mathbf{p}^0,\mathbf{q}^0)$ is called the value of the game. The min-max theorem states, in words, that every matrix game has a value and at least one optimal strategy for each player. Thus every matrix game has a solution.

Many proofs of this important theorem exist. As it stands it is an existence theorem, that is, it assures the existence of a solution to every matrix game, but it does not tell how to find it. Methods have been worked out for finding on electronic computers solutions to very large games.

Solution of examples. Although it is not possible ere to go into the details of finding solutions to games, it is easy to verify solutions when stated.

Matching pennies. The solution to this game is for each player to choose his pure strategies with equal probabilities, that is, $p_1^0 = p_2^0 = 1/2$ and $q_1^0 = q_2^0 = 1/2$. In other words the coin-flipping strategy usually employed by players of this game is actually optimal! The value of the game is 0 and it is called a fair game.

Mississippi gambler. Here the optimal strategies are $p_1^0 = 3/4$, $p_2^0 = 1/4$, and $q_1^0 = 7/12$, $q_2^0 = 5/12$, and the value of the game is $-1/4$. In other words the passenger can expect to lose an average of a quarter each time he plays the game!

Simplified poker. Two special cases for selected values of a and b will be solved.

1. $a = 4$ and $b = 8$. The matrix of (4) becomes

	Conservative	Bluffing
Conservative	0	2
Bluffing	−1	0

which is a strictly determined game. Each player should play conservatively, and the value of the game is 0.

2. $a = 8$ and $b = 4$. The matrix of (4) becomes

	Conservative	Bluffing
Conservative	0	1
Bluffing	1	0

Here the value of the game is 1/2 meaning that the game is biased in favor of player R. Optimal strategies are that each player should bluff half the time and play conservatively half the time. In ordinary parlance, bluffing is regarded as aggressive play, but note here that, to play optimally, it is necessary to bluff half of the time.

Statistical games. One of the first applications of game theory to another field was its use in the theory of statistical decision functions by Abraham Wald, the late statistician. In such statistical games, the row player is interpreted to be the statistician and the column player "Nature." The difference between Nature and a real-life player is that there is no known bias on her part. She may sometimes choose to help the statistician and sometimes not. Wald himself chose to apply the theory of games to this problem directly and gave the solution for the statistician as the min-max strategy in the matrix game. This solution has been objected to by others as being too pessimistic, and several other solutions have been proposed.

For example, in the Doctor's Dilemma, the statistician is a doctor who is testing two drugs A and B for the treatment of a certain disease. Drug A is known to be somewhat effective for treatment and never has been associated with after effects. Drug B has been spectacularly effective in certain cases but there have been undesirable after effects that, however, cannot be definitely attributed to the drug alone. The doctor analyzed the situation to be as in tabulation (8). The doctor assigned these

Doctor's choice	Nature's choice: A better than B	Nature's choice: B better than A
Uses A rather than B	10	5
Uses B rather than A	−5	20

(8)

payoffs by associating a positive monetary outcome to early recovery from the disease and a negative outcome to undesirable effects. How shall the doctor make his choice?

To protect himself from the worst that can happen using pure strategies, he would choose the first row, because it is the row whose minimum entry is maximum. If he permits himself mixed strategies, he would use the optimal strategy that selects the first row with probability 5/6 and the second row with probability 1/6. (For instance, he could roll a die and choose drug B if and only if an ace turns up.)

For many people the use of chance devices in making important decisions such as these seems repugnant. Yet the theory of games shows that, if the goal is to maximize mathematical expectation, and the analysis of the situation is correct, then such a decision rule is optimal.

Nonzero-sum games. By far the most satisfactory part of the theory of games consists of the zero-sum two-person cases, that is, in matrix games.

Applications of the theory to such areas as economics, sociology, and political science almost invariably lead to many-person nonzero-sum games. Although no universally accepted theory has been developed to cover these games, many interesting and useful attempts have been made to deal with them.

When more than two persons are involved in a conflict situation the important feature of the game becomes the coalition structure of the game. A coalition is a group of players who band together and, in effect, act as a new player in the game. There are two extremes to be considered. One is the noncooperative game in which such coalitions are banned by some means. Equilibrium-point solutions, discussed below, provide reasonably satisfactory solutions to such games. The other extreme is that in which all the players join together in a coalition to maximize jointly their total payoff. A game in which coalitions are permitted is called a cooperative game.

In the noncooperative game, each player is solely interested in his own payoff. By an equilibrium point in such a game is meant a set of mixed strategies $s_1, \ldots, s_n$ such that for each $i = 1, \ldots, n$

$$P_i(s_1, \ldots, s_i, \ldots, s_n) \geqq P_i(s_1, \ldots, s'_i, \ldots, s_n)$$

for all strategies s'_i of player i. What this means is that no player can, by changing his own strategy and assuming that the other strategies stay fixed, improve his own payoff. By a theorem of J. Nash, every game has at least one equilibrium-point solution (commonly there are several).

For example, in the Gas War, Jones and Smith own gasoline stations on opposite sides of the street. No other stations are nearby. There are only two prices they can charge, high and low, and each day they must decide which price they will charge. They are not permitted to change prices during the course of the day. Suppose their daily gross receipts are as in tabulation (9). In each pair of

	Smith	
Jones	High	Low
High	(10,10)	(6,16)
Low	(16,6)	(7,7)

(9)

numbers enclosed in parentheses the first number is Jones' receipts and the second number is Smith's receipts.

To interpret these numbers observe that if both charge the high price, each receives \$10. If both charge the low price (gas war), their income is cut to \$7 for the day. Finally, if one charges the high price and the other low, the low-priced man can not only draw business away from his immediate competitor but also from elsewhere; and the volume of his business yields him \$16, whereas his opponent gets only \$6.

The unique equilibrium point is (7,7) as can easily be checked. In other words the gas war prevails. If either player changes his strategy, he will only cut his own income, so that this solution is stable in the sense that neither can, by himself, profitably change to another strategy.

Cooperative solutions. To most people the equilibrium-point solution to the gas-war example seems irrational, because at the (10,10) point each player is better off than at the equilibrium point. But the (10,10) point is not stable, because either player can, by switching to the low price, improve his take from 10 to 16 (assuming that his opponent does not change). It is also true that at the (6,16) or (16,6) points the total amount the players take in is greater than the total amount they take in at the (10,10) point, and so either of these points is preferable. Yet, again, neither of these points is stable in the above sense.

The von Neumann-Morgenstern cooperative solution to the gas-war game is that the players should join together and cooperate to maximize their total intake; they should divide it in such a way that each player gets at least 7, the amount he can get in the noncooperative solution. In the example, each player would get 7 plus some part of the surplus profit of 8, but the theory does not indicate further how the division of the profit should be made.

J. Nash and H. Raiffa have given rationale that indicates that the surplus should be divided equally and the net payoff to each player should be \$11. The division of the surplus could be carried out by means of a payment of one player to the other or else by having Jones charge high and Smith low on alternate days, and on the other days let the reverse be true. Then their average payment will be \$11 each.

M. Shubik and G. L. Thompson have proposed another solution to the game. Suppose that Jones is richer than Smith and suppose that at the (7,7) point both players actually lose money. Because the game is played over and over again every day, one could imagine Jones making a threat to Smith as follows: "If you do not join with me in charging the high price, the day after you lower your price I shall lower mine and keep it down until you are bankrupt." Because Jones has more capital than Smith, the threat is effective and, if carried out, would result in the closing of Smith's business. If Smith tried to make a similar threat, he would simply ruin himself, and his threat is therefore suicidal Here Jones is able to enforce the (10,10) point and has the dominant position, in what otherwise appears to be a symmetric game, simply by being richer.

Simple games. An important class of n-person games for application to political behavior are the so-called simple games. Each coalition in such a game can be either winning, losing, or blocking. For instance, a winning coalition may be a set of voters who can elect their candidate, or a gro of lawmakers who can pass their bill. The pl not in a given winning coalition form a losi

alition. Finally, a coalition is blocking if neither it nor the players not in it can enforce their wishes.

It is a common observation that the power of an individual in such a voting situation is not always proportional to the number of votes he controls. This results in part from personality differences, and in part from the voting structure itself. The power index, proposed by L. S. Shapley, assigns to each player an a priori number that indicates the power that the voting structure gives to that player. It is illustrated in the following example.

Green, Jones, and Smith are on a committee. On any issue Green can cast 1 vote, Jones 2 votes, and Smith 3 votes. If it takes 4 votes to carry a measure, what are the winning, losing, and blocking coalitions and what is the relative voting power of each man?

The winning coalitions are (Green, Smith), (Jones, Smith), and (Green, Jones, Smith). Losing coalitions are (Green) and (Jones). Blocking coalitions are (Green, Jones) and (Smith).

Suppose that the men enter the committee room singly. A man is pivotal when his entry into the room changes the set of people in the room from a losing or blocking coalition into a winning coalition. There are 6 orders in which they can enter the room. The pivotal man's name is listed in boldface.

Green, Jones, **Smith**
Green, **Smith**, Jones
Jones, Green, **Smith**
Jones, **Smith**, Green
Smith, **Green**, Jones
Smith, **Jones**, Green

Since Smith is pivotal 4 out of 6 times he is assigned power $4/6 = 2/3$. Similarly, Green and Jones are pivotal once each and each are assigned power $1/6$. Note that the sum of the powers adds up to 1. Note also that Smith controls only one-half of the votes but has two-thirds of the power. Also Jones controls $1/3$ of the votes but has only $1/6$ of the power.

This model can be used to compare the actual power of a voting member with his index as given above. Thus personality factors can be separated from voting structure factors by means of the model. Generalizations of the model have also been made.

Continuous games. If, in the normalized form of the game, each player is permitted to have a continuous range of pure strategies and the payoff function is permitted to be a function of the two real variables that range over each player's strategies, the result is a continuous game. The model most frequently used is to let player R choose a number x between 0 and 1 and C choose a number y between 0 and 1, and let the payoff function $f(x,y)$ be defined on the unit square. Such games are frequently labeled "games on the square" for obvious reasons. There exist such games without solutions, but with some reasonable assumption as to properties of $f(x,y)$, it can be shown that the min-max theorem holds for these games. The same theorem has been generalized still further by permitting the strategy spaces of the players to be the more general spaces of analysis.

Machines that play games. One of the first applications of large electronic computers to numerical problems was in solving large matrix games. Several methods have been devised for finding such solutions. Currently, the most popular such method is the simplex method, devised by George Dantzig. It has been programmed on currently available machines to solve games with $m + n \leqq 250$. The principal application of the method is to solve linear programming problems, which can be shown to be equivalent to matrix games.

Computers have also been programmed to play board games such as checkers and chess. Strictly speaking they do not use the theory of games at all at present, but instead use some of the game sense of the people who devised the codes. Computers are not yet as good at playing these games as moderately good human players. Still there is hope that in the near future they will become better than human players and even eventually help to answer the game-theoretic question of whether chess must end in a draw or whether either black or white has an optimal winning strategy.

Because applications of game theory frequently lead to huge matrix games, it is clear that the future applications of the theory will rest, in part, on the ability of computers to provide the solutions to such large games.

Conclusion. The framework of game theory provides a basis on which to analyze many problems that have long been considered by philosophers and behavioral scientists. For instance, the concept of the welfare state, the theory of monopoly, the idea of the maximum good for the maximum number, and ologipolistic competition all have received renewed and sharper examination in its light. The idea of maximum good for the maximum number has been shown to be unobtainable (a min-max principle is workable). In sociology and political science, models are being developed that will help the analysis of old problems and sharpen that of the new. The voting models examined above are examples of such. *See* Information theory; Matrix theory; Probability; Stochastic process.

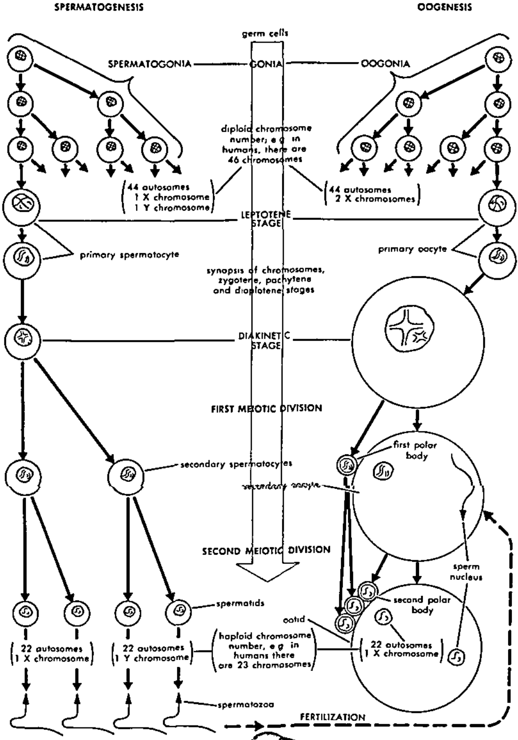
[C.L.T.]

Bibliography: *Contributions to the Theory of Games*, in Annals of Mathematics Studies, vols. 1–4, 1950–1959; J. G. Kemeny, J. L. Snell, and G. L. Thompson, *Introduction to Finite Mathematics*, 1957; T. C. Koopmans (ed.), *Activity Analysis* [illegible]

J. C. C. McKinsey, *Introduction to the Theory of Games*, 1952; O. Morgenstern and J. von Neumann, *Theory of Games and Economic Behavior*, rev. ed., 1953; M. Shubik (ed.), *Readings in Game Theory and Political Behavior*, 1954; R. M. Thrall, C. H. Coombs, and R. L. Davis (eds.), *Decision Processes*, 1954; S. Vajda, *The Theory of Games and Linear Programming*, 1956; J. D. Williams, *The Compleat Strategyst*, 1954.

Gametogenesis

The formation of the gametes, either eggs or sperm. This process involves changes in the nucleus of the preceding germ cells, or gonia, that are fundamentally alike in all animals. Common to both sexes is the occurrence, in each transforming gonium, of two meiotic, or reductional, divisions whereby the original double (diploid) set of chromosomes, characteristic of all the somatic cells of the body, is reduced to a single (haploid) set. In this process (see illustration) the chromosomes first assume the form of thin threads (leptotene stage) and each from one set joins in precise lateral apposition



Diagrammatic scheme of gametogenesis in human ma

(synapsis) with its homologue in the other set. At about this time each chromosomal thread also becomes visibly double, so that the nucleus is now composed of the haploid number of four-stranded chromosomal bodies, called tetrads. At this time, too, homologous threads, or chromatids, may exchange equivalent parts, a phenomenon known as crossing-over (*see* GENETICS). By two ensuing meiotic cell divisions, a chromatid from each tetrad goes to each of the four daughter cells, which form the spermatozoa in the male, or the egg and three polar bodies in the female. *See* OOGENESIS; SPERMATOGENESIS.

In plants, the meiotic divisions do not give rise directly to gametes but to haploid, asexual spores called tetraspores which, by mitotic divisions, form the gametophyte, some of whose cells become the gametes. The meiotic process is, however, fundamentally like that in animals. [A.TY.]

Gamma function

A particular mathematical function that can be used to express many definite integrals. There are, however, no significant applications where the gamma function by itself constitutes the essence of the solution. Instead it occurs usually in connection with other functions such as Bessel functions and hypergeometric functions.

A special case of the gamma function is the factorial $n! = 1 \cdot 2 \cdot 3 \cdot \cdots n$ (for example $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$). It is defined only for integral positive values of n. The factorial occurs for instance in the expansion

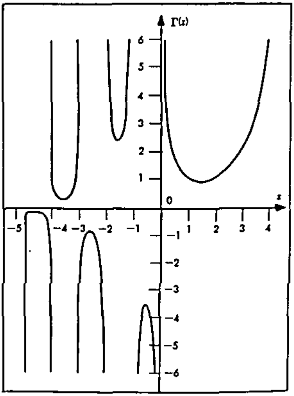
$$\exp z = 1 + z/1! + z^2/2! + z^3/3! + \cdots$$

The binomial coefficient $\binom{N}{n}$ can be expressed in terms of factorials as $N!/[n!(N-n)!]$. Many occurrences of the factorial are found in combinatorial theory (for instance, $n!$ is the number of permutations of n different elements), in probability theory, and in the applications of this theory to statistical mechanics. For large values of n, the factorial can be easily, although only approximately, computed with Stirling's formula

$$n! \approx (n/e)^n (2\pi n)^{1/2}$$

This formula is in error by a factor, which lies, for all $n \geqq 1$, between 1 and $1/(11n)$. For $n = 10$, Stirling's formula gives approximately 3598695.6 . . . instead of the accurate value $10! = 3628800$.

The gamma function can be considered as a certain interpolation of the factorial to nonintegral values of n. It is defined by the definite integral $\Gamma(z) = \int_0^\infty u^{z-1}e^{-u}\,du$. This definition is applicable to all complex values of z with positive real part. The value of $\Gamma(n+1)$ coincides with $n!$ if $n = 1, 2, 3, \ldots$. The definition of $\Gamma(z)$ can be extended to all complex values of z by the principle of analytic continuation. Other notations are in use: $z!$ or $\Pi(z)$ for $\Gamma(z+1)$. Sometimes the phrase factorial function is used for $z!$ or $\Pi(z)$ when z is arbitrary.



Representation of the gamma function $\Gamma(z)$ for real values of z.

From its definition follow the fundamental properties of the gamma function:

1. $\Gamma(z+1) = z\Gamma(z)$ (difference equation or functional equation).
2. $\Gamma(z)\Gamma(1-z) = \pi/\sin(\pi z)$,
   $\Gamma(\frac{1}{2}+z)\Gamma(\frac{1}{2}-z) = \pi/\cos(\pi z)$.
3. $\Gamma(z)\Gamma(z+\frac{1}{2}) = \pi^{1/2}2^{1-2z}\Gamma(2z)$ (Legendre's duplication formula).
4. $\Gamma(z)$ is a regular function in the complex z plane except for the points $z = -m$ ($m = 0, 1, 2, \ldots$), where it goes to infinity, as does $[(-1)^m m!(z+m)]^{-1}$, if z is close to $-m$.
5. $\Gamma(z)$ is real for real values of z, $\Gamma(1) = 0! = 1$, $\Gamma(\frac{1}{2}) = \pi^{1/2}$.

In the angular domain $-\pi < \arg z < \pi$, there exists an expansion of the gamma function, namely

$$\Gamma(z+1) = z^{z-1/2}e^{-z}(2\pi)^{1/2}\left[1 + \frac{1}{12z} + \frac{1}{288z^2} + \cdots\right]$$

which is not convergent but nevertheless useful for the numerical computation of $\Gamma(z)$ for large z. It gives approximate numerical values of the gamma function for large values of z if one takes only a finite number of terms into account. In particular for real positive values of z, the error is less than the last term taken into account. This expansion generalizes Stirling's formula.

Psi function. The logarithmic derivative of the gamma function $d\ln\Gamma(z)/dz = \Gamma'(z)/\Gamma(z)$, is de-

noted by $\psi(z)$. This function has the properties

$$\psi(z+1) - \psi(z) = \frac{1}{z}$$

$$\psi(1) = -C = -0.5772157\ldots$$

$$\psi(n+1) = -C + 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \quad \text{for } n = 1, 2, 3, \ldots$$

$$\psi(\tfrac{1}{2}) = -C - 2\ln 2$$

$$\psi(z) = -C + \sum_{n=0}^{\infty}\left(\frac{1}{1+n} - \frac{1}{z+n}\right)$$

Beta function. The beta function generalizes the binomial coefficient $\binom{N}{n}$ to noninteger values of N and n. The beta function $B(x,y)$ is by definition equal to $\Gamma(x)\Gamma(y)/\Gamma(x+y)$. Therefore one has $\binom{N}{n}^{-1} = (N+1)B(n+1, N-n+1)$.

Some of the definite integrals, whose values can be expressed by the gamma function, the psi function, and the beta function are

$$\int_0^\infty \exp(-at^n)t^z\,dt = (z+1)^{-1}a^{-(z+1)/n}\Gamma\left(\frac{z+n+1}{n}\right) \quad (\operatorname{Re} a > 0, \operatorname{Re} z > -1)$$

$$\psi(z) = \int_0^\infty \left[\frac{e^{-t}}{t} - \frac{e^{-zt}}{1-e^{-t}}\right] dt = -C + \int_0^1 \frac{1-t^{z-1}}{1-t}\,dt \quad (\operatorname{Re} z > 0)$$

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt = 2\int_0^{\pi/2} \sin^{2x-1}\zeta \cos^{2y-1}\zeta\, d\zeta \quad (\operatorname{Re} x > 0, \operatorname{Re} y > 0)$$

See BESSEL FUNCTIONS. [J.ME.]

Bibliography: A. Erdelyi (ed.), *Higher Transcendental Functions*, vol. 1, 1953; W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Special Functions of Mathematical Physics*, 1949; E. T. Whittaker and G. N. Watson, *Course of Modern Analysis*, 4th ed., 1952.

Gamma globulin

Any of the group of plasma proteins migrating most slowly (that is, having the smallest net charge) during electrophoresis by conventional methods, generally carried out at pH 8.6. While the gamma globulins have many properties in common, it is well established that in man and most mammals the

molecular weight of 156,000, but the remaining 10% have a sedimentation constant of 19 Svedberg units and a molecular weight in the neighborhood of 1,000,000. Further evidence of physical heterogeneity has been obtained by electrophoresis convection and by column chromatography. End-group analysis has revealed greater or lesser degrees of chemical heterogeneity in most species studied. An outstanding exception is the rabbit, in which all of the proteins in the normal gamma globulin fraction, as well as specific antibody globulins isolated from immunized rabbits, appear to have identical end-groups and N-terminal sequences. *See* CHROMATOGRAPHY; ELECTROPHORESIS; ULTRACENTRIFUGE.

Antibodies. The gamma globulins are of particular biological interest because the great majority of the antibody proteins migrate with this fraction. Animals injected repeatedly with an antigenic substance show striking increases in plasma concentrations of gamma globulins. However, the addi-

have been found in the β- and in the α-globulin regions but these are exceptional. Animals reared under germfree conditions and thus not exposed to stimulation by bacterial antigens have levels of gamma globulins only about one-third those of normal animals. *See* ANTIBODY; ANTIGEN; BLOOD GROUPS; GERMFREE VERTEBRATE.

The importance of the antibody function of this group of proteins in man is dramatically illustrated in congenital agamma-globulinemia, a disease characterized by gamma globulin levels below 50 mg/100 ml (normal values 600–1100 mg/100 ml). Children with this disorder suffer repeated severe bacterial infections and fail to develop circulating antibodies in response either to natural infections or to attempted vaccination. The disease is associated with a deficiency of lymphoid tissue, in particular plasma cells, which are believed to be the primary site of synthesis of the antibody globulins. Patients with this disease can be effectively protected by repeated and continuing injections of gamma globulin isolated from normal human plasma; this is further evidence that their susceptibility to infection is directly attributable to deficient antibody synthesis.

Antibody-antigen complex. The interaction between an antibody globulin and the antigenic protein stimulating its production is remarkably specific, although the specificity is by no means absolute. For example, analogous proteins taken from different species (such as serum albumins) will generally show qualitative cross reactions. The specificity is of such a degree, however, as to make immunochemical methods extremely valuable for identification and quantification of small amounts of protein in a mixture. *See* BIOLOGICAL SPECIFICITY; IMMUNITY.

The antibody-antigen interaction is of great interest to the protein chemist as an example of a highly specific protein-protein interaction. By modifying the antigen or the antibody chemically or physically it has been shown that not all of t

structural features of the protein molecule are essential to the formation of antibody-antigen complexes. Certain modifications of antibody protein, in particular partial proteolytic degradation, prevent formation of immune precipitates but do not prevent complex formation. These partial antibodies, added to a solution of antigen, prevent precipitation when unmodified antibodies are subsequently added. Such antibodies are termed blocking antibodies. Antibodies with this blocking activity are also found in some untreated immune serums. One hypothesis consistent with these observations is that in precipitin reactions antibody and antigen combine at two or more points and can therefore readily form large aggregates, whereas incomplete or blocking antibodies can only combine at one site, yielding a soluble complex. *See* PRECIPITIN TEST.

Mechanism of antibody synthesis. The exact mechanism of specific antibody synthesis is neither known nor likely to be established until the general process of protein synthesis is understood. Two current hypotheses may be mentioned. According to the Breinl-Haurowitz-Mudd theory, the actual sequence of steps in the synthesis of globulin protein from amino acids is altered by the presence of antigen protein in the cell. Thus the globulin molecules destined to be antibody globulins acquire a different amino-acid sequence and a different three-dimensional configuration, making their structure complementary to that of the antigen. According to the Pauling theory, on the other hand, the antibody globulins are all synthesized in the same fashion as are normal gamma globulins, except that in the presence of antigen molecules two areas (ends) of the polypeptide chain take on a configuration complementary to that of the antigen. In favor of the latter theory is the finding that the N-terminal sequence of antibody globulins in the rabbit is the same as that of normal rabbit gamma globulin. The sequence in the rest of the molecule, however, has not been determined. Further studies of this kind, relating the detailed structure of antibody globulins to that of normally circulating gamma globulins, will be necessary before any final conclusion can be reached. [D.ST.]

Gamma rays

Electromagnetic radiation emitted from certain atomic nuclei as a result of rearrangements within the nuclei leading to a lower energy content. For the theory of gamma emission, *see* RADIOACTIVITY. *See also* ELECTROMAGNETIC RADIATION.

The energies of nuclear gamma rays (γ-rays) range from practically zero up to the order of 10 Mev. The frequency ν of a γ-ray (or any photon) depends on its energy E according to $E = h\nu$, where h is Planck's constant. A photon carries the momentum $p = h\nu/c$, where c is the velocity of light. *See* PHOTON.

Interaction with matter. The three types of interaction with matter which together are responsible for the observable absorption of γ-rays, namely,

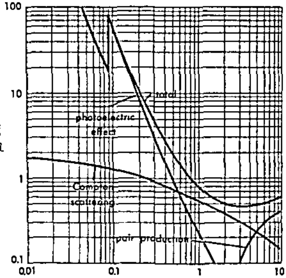


Fig. 1. Partial and total attenuation coefficients for lead as a function of gamma energy. (*National Bureau of Standards*)

Compton scattering, the photoelectric effect, and pair production, are discussed in detail elsewhere. *See* COMPTON EFFECT; PAIR PRODUCTION (ELECTRON-POSITRON); PHOTOEMISSION.

The energy of a photon may be absorbed totally or partially in interaction with matter; in the latter case, the frequency of the photon is reduced and its direction of motion is changed. Photons are thus absorbed not gradually, but rather in discrete events, and one interaction is sufficient to remove a photon from a collimated beam of γ-rays. The intensity I of a beam decreases exponentially

$$I = I_0 e^{-\mu x}$$

where x is the path length, I_0 is the initial intensity and μ is the linear attenuation coefficient, which is characteristic of the material and the gamma energy (Fig. 1).

Gamma detectors. Gamma ray counters, devices used for detecting γ-radiation, either yield information about the integrated intensity within a time interval or register each photon separately. The working principle of such counters must always be the interaction with matter, primarily the production of fast electrons by γ-quanta, and thus any detector for electrons will also detect γ-rays. The efficiency, that is, the probability that a photon passing through the sensitive volume (the volume where an event can be registered) will be detected, is generally poor. Ionization chambers are used mainly as integrating instruments. Because of its simplicity, the Geiger tube is frequently used for general purposes. Above 100 kev its efficiency is quite low; the efficiency is less than 1% at 1 Mev. *See* GEIGER-MÜLLER COUNTER; IONIZATION CHAMBER.

The most generally used gamma counter, the NaI(Tl) or sodium iodide crystal scintillation counter, is treated in the next section.

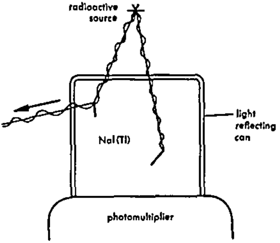


Fig. 2. Illustration of γ-ray absorption in a sodium iodide crystal. In the event at the left, the photon coming from the source undergoes Compton scattering in the crystal and the secondary quantum escapes unabsorbed. To the right a photon is fully absorbed because of the photoelectric effect. In both situations, the electrons are completely stopped in the crystal.

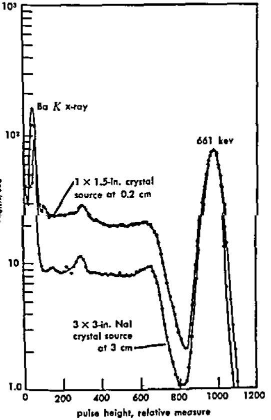


Fig. 3. Scintillation spectrum of Cs^{137}, emitting only one nuclear γ-ray of energy $h\nu = 661$ kev. The Ba x-ray is emitted after internal conversion in the source. (From K. Siegbahn, ed., *Beta- and Gamma-Ray Spectroscopy*, North Holland, 1955)

Energy measurement. Gamma ray spectrometers are used for determining the energy of γ-rays, or the energy composition, if several energies are present simultaneously. The result ideally takes the form of a curve showing the number of photons as a function of their energy. The discrete energies emitted by nuclei are in practice observed as peaks of finite width. Figures of merit for such instruments are the relative line width, which is a measure of the resolving power, and the efficiency.

Scintillation methods. The simplest and most sensitive spectrometer is the NaI(Tl) crystal, which is a special type of a scintillation counter. The thallium (Tl) is a sensitizer (*see* CRYSTAL COUNTER; SCINTILLATION DETECTOR, LIQUID). The NaI(Tl) crystal differs from organic crystals in its much higher absorption coefficient. With sufficiently large crystals, the counting efficiency may be practically 100% at a few hundred kev and more than 50% at higher energies. An incoming photon may be totally or partially absorbed (see Fig. 2). The total absorption of photons results in a pulse of light whose amplitude is proportional to the γ-energy, whereas partially absorbed photons create a continous distribution of lower pulse heights (Fig. 3). The *total absorption peaks (photo peaks) are narrow* enough to permit energies to be measured with an error of less than 1%.

Crystal spectrometer. The crystal diffraction spectrometer, also used for the investigation of x-rays, has found use with γ-rays up to 1 Mev energy.

Beta ray spectrometers. The most generally applied methods involve the use of β-ray spectrometers to measure the energies of internal conversion electrons, photoelectrons, and Compton electrons produced by γ-rays. For an extended discussion of β-ray spectrometers, *see* BETA RAYS. [C.B.]

Bibliography: R. D. Evans, *The Atomic Nucleus*, 1955; W. J. Price, *Nuclear Radiation Detection*, 1958; K. Siegbahn (ed.), *Beta- and Gamma-Ray Spectroscopy*, 1955.

Gammaridea

A suborder of Amphipoda. These crustaceans are commonly known as scuds in aquatic environments, and as sandhoppers on beaches. In 1956, the suborder comprised 3200 species in 672 genera and 57 families. The majority of the species are marine but more than 500 species are limnetic, to an altitude of 4000 meters, or are subterranean; 80 species are terrestrial (family Talitridae). The terrestrial species are confined mainly to beaches in high latitudes but in the Indo-Pacific insular region they penetrate far inland and to an altitude of 3000 meters in moist environments.

Gammarideans are one of the common medium-sized crustacean groups, 3–12 mm long, found in intertidal regions, in algal zones, and on the bottom of the sea along continental margins. They penetrate to the greatest ocean depths. The largest species, *Alicella gigantea*, 140 mm in length, occurs in the deep sea. Pelagic Gammaridea are less abun-

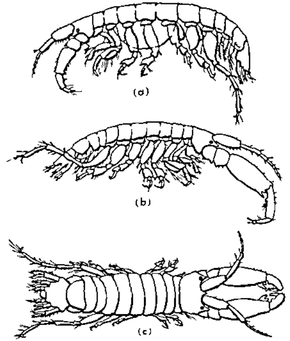


Fig. 1. Corophiidae. *Corophium crassicorne*, Bruzelius, a tube-building amphipod: (a) Female, lateral view. (b) Male, lateral view. (c) Male, dorsal view. (From G. O. Sars, *An Account of the Crustacea of Norway*, vol 1, 1895)

dant and less diversified than the pelagic Hyperiidea and are best represented by members of the Lysianassidae and Eusiridae.

Except for two species of Ingolfiellidea, all freshwater amphipods are gammarids, mainly in the families Gammaridae and Hyalellidae. Long isolation of Gammaridae in Lake Baikal has resulted in an endemic population of more than 200 species. To a smaller degree, this same kind of adaptive radiation has occurred in the Caspian Sea and among the Hyalellidae in Lake Titicaca. *See* Biotic isolation; Speciation.

Morphology. Gammaridea are usually compressed laterally and are poor walkers, unlike the Isopoda, except for the families Corophiidae and Cheluridae, which have depressed bodies. The head and segments are free, lacking a carapace. The head bears two pairs of antennae and six kinds of mouthparts, some of them paired. The mandibles usually bear cutting and trituration surfaces for chewing but a few species lack these devices or have the mouthparts modified for sucking the tissues of animals such as compound ascidians. The posterior pair of mouthparts, the maxillipeds, are considered to belong to a thoracic segment fused to the head, so that some carcinologists number the free thoracic segments as 2–8. However, practicing taxonomists number them 1–7 for descriptive simplicity.

Each thoracic segment bears a pair of appendages. The first two pairs are usually prehensile and are known as gnathopods but their function is not only the manipulation of sediments and food, but also the copulatory amplexion of the female by the male. Thus, the male gnathopods, or gamopods, are usually better developed than those of the female. The remaining five thoracic segments each bear a pair of pereiopods, or walking legs, arranged in two groups.

The abdomen is composed of six segments, with each of the first three bearing a pair of biramous pleopods or swimmerets, and each of the last three bearing a pair of uropods, semirigid backward projecting appendages. A flap, the telson, is attached to the sixth segment and is often lobed.

Internal structures. The simple mouth and esophagus lead into a stomach occupying part of the head and the anterior thorax. The food is triturated in the stomach by a gastric mill composed of two setose plates. Two pairs of tubular hepatic ceca open into the stomach, the tubes extending posteriorly, often into the abdomen. The rest of the gut is a simple tube, distinguishable into regions histologically and muscularly. A pair of rectal ceca, debatedly excretory, opens into the posterior gut. The anal opening is on the last abdominal segment. The main circulatory mechanism is in the thorax, unlike the Isopoda where it is abdominal. This contrast is related to the differences in respiration between the two orders. In Isopoda the abdominal pleopods serve as branchiae, while in Amphipoda five or six pairs of separate branchial lamellae are attached to the medial bases of the thoracic legs. The heart is a tube with three pairs of lateral openings from a surrounding sinus, the pericardium. Blood from the pericardium enters the heart and is pumped through anterior and posterior aortas to smaller vessels serving the head, body, and appendages. The blood, which is colorless, is returned to the pericardium through spacious sinuses.

The nervous system is composed of two dorsal cephalic ganglia connected by circumesophagic commissures to a ventral chain of 12 ganglia, including 1 subesophageal, 7 thoracic and 4 abdominal. The ganglia are connected by double commissures.

The sense organs consist of eyes and small cuticular organs such as setae, calceoli of debatable function, and sensory filaments on the antennae, especially the male. In male Phoxocephalidae, the eyes are greatly enlarged, which is probably related to their habit of leaving their mud burrows and swimming to the sea surface at night. Corneal lenses (cuticular) are developed in the Ampeliscidae, while the remaining families possess ommatidea only. Deep-sea species usually lack eyes, and statocysts are unknown in the Gammaridea.

A pair of excretory organs, the antennal glands, discharges on the second segments of the second antennae. A similar pair of ducts, of unknown origin, opens on the lower lip.

Reproduction. The sexes are separate, and the reproductive systems are simple tubular gonads and ducts, discharging ventrally on the fifth thoracic segment in the female and the seventh in the male. Copulatory amplexion of the sexes occurs while

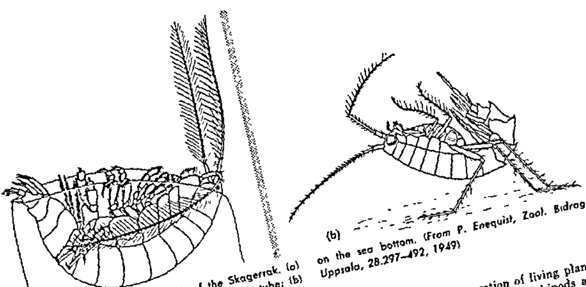


Fig. 2. Soft-bottom amphipods of the Skagerrak. (a) *Haploops tubicola*, in feeding position in its tube; (b) *Melphidippella macra*, in its inverted walking position on the sea bottom. (From P. Enequist, Zool. Bidrag Uppsala, 28.297–492, 1949)

swimming and at the time of molting of the female, when the reproductive orifices are soft and expandable enough to permit passage of the eggs. External fertilization occurs, and the eggs are held in the female brood pouch which is composed of four pairs of setose plates attached ventrally to thoracic segments 3–6. Development generally requires 12–30 days. The young then hatch as miniature adults and leave the brood pouch hours to days later. The major postembryonic changes concern the secondary sexual characteristics, such as the development of brood lamellae in the female, and the differential enlargement of the gnathopod pairs, especially in the male. Males often develop larger or more sense organs than do females.

The growth of amphipods is poorly known, except that size increases occur as part of the ecdysial mechanism characteristic of arthropods in general. One species, *Gammarus chevreuxi*, is known to reach sexual maturity after six molts, the female then producing a brood of young during each succeeding instar of growth, until death at about the twelfth instar. The length of life has been recorded as 1–2 years or less. The number of eggs produced varies from one or two to more than 200. This depends on the species and increases with the size or age of the adults.

Systematics. The basic morphologic plan in the Gammaridea is rather stable, unlike the Isopoda, for instance, where many morphological types have developed. Thus, systematic partition of amphipods is based on minor characters, usually the mouthparts. For this reason, specific identification of amphipods is difficult, except by specialists.

Habits. Gammaridea are largely scavengers, feeding on organic debris or detritus which falls to the ocean bottom. In many species the feeding is selective, whereas in others indiscriminate feeding on mud containing small organic particles occurs. The well-developed chewing mouthparts of most amphipods permit the maceration of living plants, such as algae or aquatic grasses. Amphipods also feed on large dead marine animals. This knowledge has been utilized by Alaskan hunters to clean mammalian skulls by immersing them in wire cages in the sea. Swarms of amphipods are attracted to the meaty skull and are known to clean it within a day. Similar usefulness is known in the dermestid beetles.

Species in the families Ampeliscidae, Photidae, and Corophiidae build nesting tubes attached to solid intertidal objects or lying on the sea bottom. The tubes are spun either from secreting glands on the last two pairs of pereiopods or from cuticular glands on the body. Occasionally, masses of the tubes foul ship bottoms. These animals use their well-developed antennae for straining food particles.

Species in the Phoxocephalidae and Haustoriidae have strongly spinose appendages for burrowing into bottom sediments. Some of these ingest mud, while others are selective deposit feeders.

Semiparasitic and commensal species with sucking or lapping mouthparts are known in the families Stenothoidae, Leucothoidae, and Dexaminidae. They inhabit coelenterates, ascidians, sponges, or grasp lobsters and fish. However, no amphipods are as degenerate as some species of parasitic isopods and copepods. See AMPHIPODA [J.L.B.]

Ganglion

A group of nerve cell bodies, usually located outside the brain or spinal cord. A ganglion which is located inside the cord is called a nucleus.

The dorsal root ganglia are rounded clusters of cell bodies and fibers, surrounded by a connective tissue covering located on the dorsal, or se[illegible] roots of each spinal nerve. They lie just ou[illegible] cord and contain the cell bodies of sens[illegible] fibers originate in some [illegible]

receptor. Stimulation of such a receptor (pain, temperature, or touch) causes a nerve impulse to pass into the dorsal ganglion where a synapse is made with secondary fibers that go to higher levels. *See* SENSATION; SENSE ORGAN.

Other ganglia are given specific names which indicate their function or location, such as acoustic, cardiac, carotid, jugular, celiac, and sympathetic ganglia. Sympathetic ganglia, on either side of the vertebral column, unite by fiber strands to form a sympathetic chain. Preganglionic fibers run along this path until they reach their terminal point. The relay to postganglionic neurons occurs in these ganglia.

The term ganglion may also apply to a tumorlike, often cystic, growth found on tendons, joints, and other connective tissues. *See* BRAIN; SPINAL CORD; TUMOR. [E.C.ST.]

Gangrene

A form of tissue death, or necrosis, usually occurring in an extremity and due to insufficient blood supply.

If no bacterial contamination is present, the part becomes dry, greenish-yellow, and finally turns brown or black. This is known as mummification. A sharp inflammatory border marks the edge of the adjacent viable tissue. This dry gangrene is seen most often in small portions of the extremities, like the fingers and toes. Senile gangrene is the form caused by deterioration of blood supply in the elderly, usually as the result of progressive arteriosclerosis. Similar types are often present in diabetes, Reynaud's disease, and Buerger's disease (thromboangiitis obliterans).

When bacterial infection intervenes, putrefaction ensues, thus producing the moist or wet type of gangrene. Occasionally the invading microorganisms are gas-producing species of the genus *Clostridium*, and rapidly progressive gas gangrene is the result. *See* BACILLACEAE.

Moist gangrene may occur anywhere in the body that the blood supply is blocked and bacterial contamination occurs. The size of the area varies from a small, asymptomatic, local lesion to the destruction of an entire organ, like the gallbladder or appendix. The parts most commonly affected by dry gangrene may also become contaminated, with the subsequent production of a wet gangrene. *See* GANGRENE, GAS. [E.C.ST.]

Gangrene, gas

A localized, but rapidly spreading, necrotizing wound infection, characterized by extensive edema with gas production and discoloration of the tissue, and often accompanied by a putrefactive odor. The disease commonly arises following dirt contamination of deep wounds or septic abortion. The microbial flora of the gangrenous wound usually comprises one or more species of toxigenic anaerobic bacteria mixed with nontoxigenic anaerobic species, aerobic species, or both. The following gram-positive, sporeforming species are generally considered as the primary causal agents of gas gangrene: *Clostridium perfringens* (*Cl. welchii*), *Cl. septicum* (*Vibrion septique*), *Cl. novyi* (*Cl. oedematiens*) and *Cl. bifermentans*. Anaerobic streptococci may also cause gas gangrene. *See* BACILLACEAE; LACTOBACILLACEAE.

Gangrene does not necessarily follow the presence of these organisms in a wound, as initiation of the disease depends on the virulence of the organisms and other factors relating to the resistance of the host. The virulence of the gangrene-producing organisms depends on the toxins produced. Some of the toxins have been identified as specific enzymes such as hyaluronidase, collagenase, and lecithinase.

The rapidly spreading nature of the disease precludes extensive laboratory diagnostic aids because therapeutic measures, possibly including amputation, usually must be instituted before laboratory results are available. Procedures should be begun, however, for isolation of all species present and determination of toxin type and drug sensitivity. Initial thorough surgical debridement (removal of foreign material and devitalized tissue) of all contaminated wounds is essential. Polyvalent antitoxins are available for prophylactic and therapeutic use, and penicillin and tetracycline antibiotics are valuable in some cases. Formol toxoids and alum-precipitated formol toxoids for active immunization against the more important species have been produced but are not routinely used. The effectiveness of antibiotic therapy seems to depend on the species involved in the infection and the elapsed time between injury and treatment. *See* ANTIGEN; BIOLOGICALS; ENZYME; GRAM'S STAIN; TOXIN, BACTERIAL; TOXIN-ANTITOXIN REACTION. [L.S.M.]

Bibliography: L. D. S. Smith, *Introduction to the Pathogenic Anaerobes*, 1955.

Gannet

Any of three species of large, gull-like, fish-eating marine birds of the genus *Moris*, family Sulidae, order Pelecaniformes. One species, the northern gannet, *M. bassana*, occurs along the Atlantic Coast of the United States. Adults are white, with

The gannet, *Moris bassana*; length to 40 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

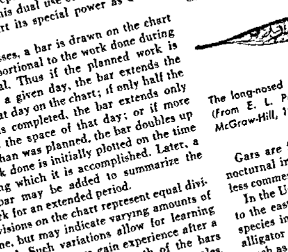
the terminal two-fifths of the wing black. Immature birds are dark brown, spotted with white. The gannet is larger than a goose; it has a long, pointed tail, and carries its strong, straight bill downward when in flight. *See* PELECANIFORMES. [J.D.B.]

Gantt chart

In production planning and control, a bar chart depicts the work planned and done in relation to time. The Gantt chart differs significantly from conventional bar graphs. A division of space on the chart represents both a time interval (hour, day, week, as appropriate) and the amount of work to be done during that interval (machine pieces, worker production, department output, or whatever is being plotted). This dual use of the chart space gives the Gantt chart its special power as a management tool.

As work progresses, a bar is drawn on the chart with a length proportional to the work done during each time interval. Thus if the planned work is accomplished on a given day, the bar extends the full length of that day on the chart; if only half the planned work is completed, the bar extends only halfway across the space of that day; or if more work is done than was planned, the bar doubles up on itself. Work done is initially plotted on the time interval during which it is accomplished. Later, a cumulative bar may be added to summarize the status of work for an extended period.

Equal divisions on the chart represent equal divisions of time, but may indicate varying amounts of work planned. Such variations allow for learning characteristics as operators gain experience after a product is modified. Thus the length of the bars drawn on the chart may also be to varying scales. The result is a dynamic chart depicting graphically what has happened and displaying in easily interpreted form the status of each job. The chart is used in the shop for machine loading, worker assignments, and project planning; in the sales office for order scheduling; and in the laboratory as an aid in directing research and development. By coded notes on the chart, causes for departures from the planned amount of work may be recorded. Such notes enable one to take corrective action on current jobs and to plan more accurately for future jobs. *See* INDUSTRIAL CONTROL. [P.F.CL.]

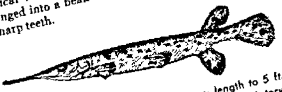


Gantt chart showing (a) daily schedule, (b) work accomplished, and (c) cumulative schedule and work. (H. B. Maynard, *Industrial Engineering Handbook*, McGraw-Hill, 1956)

Bibliography: W. Clark, *The Gantt Chart*, 3d ed., 1952.

Gar

Any of about 13 species of primitive ganoid, freshwater fishes from North and Central America and the West Indies. Gars are covered with an armor of thick, hard ganoid scales. They are slender, cylindrical fishes, with the forepart of the head prolonged into a beak and armed with several rows of sharp teeth.



The long-nosed gar, *Lepisosteus osseus*; length to 5 ft. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Gars are fierce, swift, predatory fish, primarily nocturnal in their feeding. They are almost worthless commercially.

In the United States gars are restricted primarily to the eastern part of the country. There are four species in the Mississippi Valley. The largest, the alligator gar, *Lepisosteus spatula*, which weighs as much as 300 lb, is considered a good sport fish, and is sought with tarpon tackle. *See* SEMIONOTIFORMES. [J.D.B.]

Garlic

A hardy perennial (*Allium sativum*) of Asiatic origin and belonging to the plant order Liliales. Garlic is grown for its pungent bulbs, segments of which are used primarily for seasoning. Europeans have grown it for more than 2000 years. Propagation is commonly by bulb segments, sometimes called cloves; seeds are seldom produced. Cultural practices are similar to those used for onions. Popular varieties are Italian, Tahiti, and Creole or Mexican. Harvest of the mature dry bulbs is 7–8 months after planting. Garlic salt is made from dehydrated cloves. California is the most importan producing state; smaller acreages are planted garlic in Louisiana and Texas. The total annu farm value in the United States is approximat $1,800,000. *See* LILIALES; ONION; VEGETA GROWING.

Garnet

A generic term applied . a group of silicates that are crystalli that conform to the ical

$A_3B_2(SiO_4)_3$, where A is Fe^2, Mn^2, Mg, or Ca, and B is Al, Fe^3, or Cr^3.

Garnet group. The group is divided into a number of individual mineral species on the basis of chemical composition. The names of the more common of these species and the idealized compositions to which they refer are listed.

Pyrope $Mg_3Al_2(SiO_4)_3$	Grossularite $Ca_3Al_2(SiO_4)_3$
Almandite $Fe_3Al_2(SiO_4)_3$	Andradite $Ca_3Fe_2(SiO_4)_3$
Spessartite $Mn_3Al_2(SiO_4)_3$	Uvarovite $Ca_3Cr_2(SiO_4)_3$

A wide range of solid solubility exists in the group by the mutual substitution of the several cations of the A position or of the B position. Solid solubility is particularly marked between pyrope, almandite, and spessartite, comprising the so-called Pyralspite Series, and between grossularite and andradite in the Grandite Series. Since the chemical composition of garnets as found in nature is gradational, the species name is arbitrarily assigned on the basis of the A cation and of the B cation that is dominant in atomic per cent. Many of the pure end compositions have been synthesized in the laboratory, including members of the group not found naturally.

The garnets usually are well crystallized. The typical crystal forms are the dodecahedron and the trapezohedron (211), either alone or in combination. Garnet also occurs as granular masses and as disseminated grains. Rarely, it is dense and microcrystalline, the green types then closely resembling nephrite.

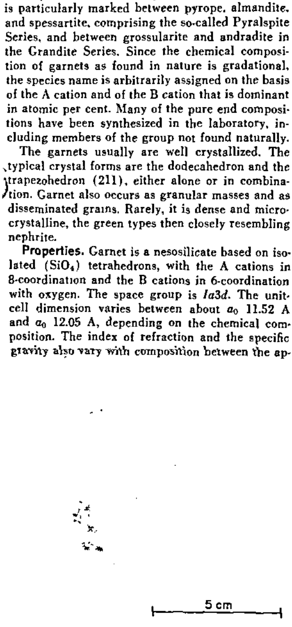
Properties. Garnet is a nesosilicate based on isolated (SiO_4) tetrahedrons, with the A cations in 8-coordination and the B cations in 6-coordination with oxygen. The space group is *Ia3d*. The unit-cell dimension varies between about a_0 11.52 A and a_0 12.05 A, depending on the chemical composition. The index of refraction and the specific gravity also vary with composition between the approximate limits 1.71–1.89 and 3.53–4.33, respectively. These properties, sometimes supplemented with qualitative chemical tests, can be used for the identification of the individual garnet species. Graphical representations of the variation in unit-cell size, specific gravity, and index of refraction with variation in chemical composition are available. *See* SILICATE MINERALS.

The color of garnet is primarily determined by the chemical composition. Common garnet, chiefly comprising almandite and in part spessartite and andradite, is generally reddish brown to black in color; grossularite usually is white, yellowish-brown to brown; uvarovite, emerald green; spessartite, dark hyacinth red, often with a tinge of violet; pyrope, deep red. When transparent and finely colored, garnet is prized as a gem stone. The gem varieties include pyrope chiefly, and some almandite and andradite. Demantoid is a rare grass-green to emerald-green gem variety of andradite. The Bohemian garnet jewelry is cut from pyrope. Asteriated varieties containing minute oriented inclusions of rutile are known. The hardness of garnet ranges from 6½ to 7½ on Mohs scale. Cleavage is lacking, although a dodecahedral parting sometimes is present, and the fracture is subconchoidal to irregular. The relatively high hardness and the angular shape of crushed grains of garnet lead to the use of the mineral as an abrasive. *See* GEM.

Normally, garnet is optically isotropic. Some species, particularly the grossularite and andradite of contact metamorphic deposits, show biaxial optical anomalies of a sectoral nature. It is relatively stable chemically but sometimes is found surficially or completely altered to chlorite or to mixtures of hornblende or biotite with feldspar.

Occurrence. Garnet is a very common mineral. Almandite occurs widely in gneissic and schistose metamorphic rocks formed by the recrystallization of argillaceous sedimentary rocks at relatively elevated temperatures and pressures. The presence or absence of garnet and of certain associated minerals in metamorphic rocks of appropriate chemical composition is an index of the grade or intensity of the metamorphic process. Grossularite is a typical mineral of contact metamorphic zones in limestone. Andradite also occurs in such deposits, particularly in association with magnetite and iron-rich silicates, and it is found as an accessory mineral in igneous rock types containing feldspathoids. Uvarovite chiefly occurs in chromite deposits. Spessartite and almandite high in manganese occur in granite pegmatites and in acidic igneous rocks, and spessartite also forms during the thermal metamorphism of argillaceous sedimentary manganese deposits. Pyrope occurs in basic and ultrabasic igneous rocks, such as peridotite, and in serpentine rocks formed from them. Garnet is one of the most persistent and widespread minerals in detrital sediments. In beach and river sands it accumulates with zircon, ilmenite, monazite, and other heavy and resistant minerals in the so-called black sands.

[C.FR.]



Garnet dodecahedron from Salida, Colorado (Brooks Museum, University of Virginia)

Garnierite

A mineral used as a minor ore of nickel. Garnierite is a monoclinic serpentine, but commonly occurs in reniform to earthy masses. The hardness is 2–3 on Mohs scale, and the specific gravity is 2.2–2.8. The luster is earthy and the color apple-green to white. The composition may be expressed by the formula $(Ni,Mg)SiO_3{\cdot}nH_2O$, but varies greatly in the relative amounts of nickel and magnesium and in the amount of water. Garnierite is a secondary mineral (a nickel-bearing serpentine) derived from the alteration of nickel-bearing olivine-rich rocks. It has been mined as an ore of nickel in New Caledonia, U.S.S.R., South Africa, and Madagascar. In the United States it is found at Riddle, Ore., and Webster, N.C. *See* NICKEL.

[C.S.HU.]

Garter snake

Any of several species of the genus *Thamnophis*, 11 of which occur in the United States. These snakes range from small to moderate size, and have heavily keeled scales. They are terrestrial, but some species are commonly found near water. Most of them are marked with a middorsal stripe and a pair of lateral stripes of yellow or orange. The body color varies from gray through brown to black, frequently marked with bright red blotches. *T. sirtalis*, the common garter snake, sometimes called the ribbon snake, is a good example of the genus. It is abundant throughout most of the eastern United States. Other forms occur westward to California and southward into Mexico. Like the genus *Natrix*, to which they are closely related, garter snakes produce large litters of living young. Their principal food appears to be frogs, tadpoles, and earthworms, although they probably eat whatever kind of animal food is available. *See* REPTILIA.

[J.D.B.]

Gas

A phase of matter characterized by relatively low density, high fluidity, and lack of rigidity. A gas expands readily to fill any containing vessel Usually a small change of pressure or temperature produces a large change in the volume of the gas. The equation of state describes the relation between the pressure, volume, and temperature of the gas. In contrast to a crystal, the molecules in a gas have no long-range order.

At sufficiently high temperatures and sufficiently low pressures, all substances obey the ideal gas or perfect gas equation of state

$$pv = RT$$

where p is the pressure, T is the absolute temperature, v is the molar volume, and R is the gas constant. The absolute temperature T expressed on the Kelvin scale is simply related to the temperature t expressed on the Centigrade scale.

$$T = t + 273.16$$

The gas constant is

$$R = 82.0567 \text{ cm}^3\text{-atm/(mole) (°K)}$$
$$= 82.0544 \text{ ml-atm/(mole) (°K)}$$

The molar volume is the molecular weight divided by the gas density.

Empirical equations of state. At lower temperatures and higher pressures, the equation of state of a real gas deviates from that of a perfect gas. Various empirical relations have been proposed to explain the behavior of real gases. The following equations are frequently used:

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT \qquad \text{J. van der Waals (1899)}$$

$$\left(p + \frac{a}{Tv^2}\right)(v - b) = RT \qquad \text{P. E. M. Berthelot (1907)}$$

$$pe^{a/vRT}(v - b) = RT \qquad \textit{F. Dieterici (1899)}$$

In these equations, a and b are constants characteristic of the particular substance under consideration. In a qualitative sense, b is the excluded volume due to the finite size of the molecules and roughly equal to four times the volume of one mole of molecules. The constant a represents the effect of the forces of attraction between the molecules. In particular, the internal energy of a van der Waals gas is $-a/v$. None of these relations gives a good representation of the compressibility of real gases over a wide range of temperature and pressure. However, they reproduce qualitatively the leading features of experimental pressure-volume-temperature surfaces.

Schematic isotherms of a real gas, or curves showing the pressure as a function of the volume for fixed values of the temperature, are shown in Fig. 1. Here T_1 is a very high temperature and its isotherm deviates only slightly from that of a perfect gas, T_2 is a somewhat lower temperature where the deviations from the perfect gas equation are quite large, and T_c is the critical temperature. The

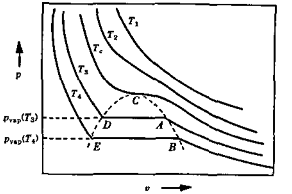


Fig. 1. Schematic isotherms of a real gas. Here C is the critical point. The points A and B give the volume of the gas in equilibrium with the liquid phase at their respective vapor pressures In a similar manner, D and E are the volumes of the liquid in equilibrium with gas phase.

critical temperature is the highest temperature at which a liquid can exist. That is, at temperatures equal to or greater than the critical temperature, the gas phase is the only phase that can exist (at equilibrium) regardless of the pressure. Along the isotherm for T_c lies the critical point, C, which is characterized by zero first and second partial derivatives of the pressure with respect to the volume.

$$(\partial p/\partial v)_c = (\partial^2 p/\partial v^2)_c = 0$$

At temperatures lower than the critical, such as T_3 or T_4, the equilibrium isotherms have a discontinuous slope at the vapor pressure. At pressures less than the vapor pressure, the substance is gaseous; at pressures greater than the vapor pressure, the substance is liquid; at the vapor pressure, the gas and liquid phases (separated by an interface) exist in equilibrium.

Along one of the isotherms of the empirical equations of state discussed above, the first and second derivatives of the pressure with respect to the volume are zero. The location of this critical point in terms of the constants a and b is shown in the table in which p_c and v_c are the pressure and volume at the critical temperature.

	Van der Waals	Berthelot	Dieterici
p_c	$\dfrac{a}{27b^2}$	$\left(\dfrac{aR}{216b^3}\right)^{1/2}$	$\dfrac{a}{4e^2b^2}$
v_c	$3b$	$3b$	$2b$
T_c	$\dfrac{8a}{27Rb}$	$\left(\dfrac{8a}{27Rb}\right)^{1/2}$	$\dfrac{a}{4Rb}$
$\dfrac{p_cv_c}{RT_c}$	0.3750	0.3750	0.2706

Some typical values of p_cv_c/RT_c for real gases are as follows: 0.30 for the noble gases, 0.27 for most of the hydrocarbons, 0.243 for ammonia, and 0.232 for water. The van der Waals and Berthelot equations of state cannot quantitatively reproduce the critical behavior of real gases because no substance has a value of p_cv_c/RT_c as large as 0.375. The Dieterici equation gives a good representation of the critical region for the light hydrocarbons but does not represent well the noble gases or water.

At temperatures lower than the critical point, the analytical equations of state, such as the van der Waals, Berthelot, or Dieterici equations, give S-shaped isotherms as shown in Fig. 2. From thermodynamic considerations, the vapor pressure is determined by the requirement that the cross-hatched area DEO be equal to the cross-hatched area AOB. Under equilibrium conditions, the portion of this isotherm lying between A and D cannot occur. However, if a gas is suddenly compressed, points along the segment AB may be realized for a short period until enough condensation nuclei form to create the liquid phase. Similarly, if a liquid is suddenly overexpanded, points along DE may occur for a short time. For low temperatures, the point E may represent a negative pressure corresponding to the tensile strength of the liquid.

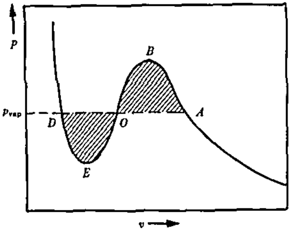


Fig. 2. Schematic low temperature isotherm as given by van der Waals, Berthelot, or Dieterici equations of state. Here the line DOA corresponds to the vapor pressure. The point A gives the volume of the gas in equilibrium with the liquid phase and D gives the volume of the liquid. The segment of the curve DE represents overexpansion of the liquid. The segment AB corresponds to supersaturation of the vapor. However, the segment EOB could not be attained experimentally.

However, the simple analytical equations of state cannot be used for a quantitative estimate of these transient phenomena. Actually, it is easy to show that the van der Waals, Berthelot, and Dieterici equations give poor representations of the liquid phase since the volume of most liquids (near their freezing point) is considerably less than the constant b.

Principle of corresponding states. In the early studies, it was observed that the equations of state of many substances are qualitatively similar and can be correlated [illegible] the variables [illegible] dimen- [illegible] r, are defined by dividing each variable by its value at the critical point.

$$p_r = p/p_c$$
$$T_r = T/T_c$$
$$v_r = v/v_c$$

In its most elementary form, the principle of corresponding states asserts that the reduced pressure, p_r, is the same function of the reduced volume and temperature, v_r and T_r, for all substances. An immediate consequence of this statement is the statement that the compressibility factor

$$z = pv/RT$$

is a universal function of the reduced pressure and the reduced temperature. This principle is the basis of the generalized compressibility chart of O. A. Hougen and K. M. Watson (Fig. 3). This chart was derived from data on the equation-of-state behavior of a number of common gases.

It follows directly from the principle of corresponding states that the compressibility factor at

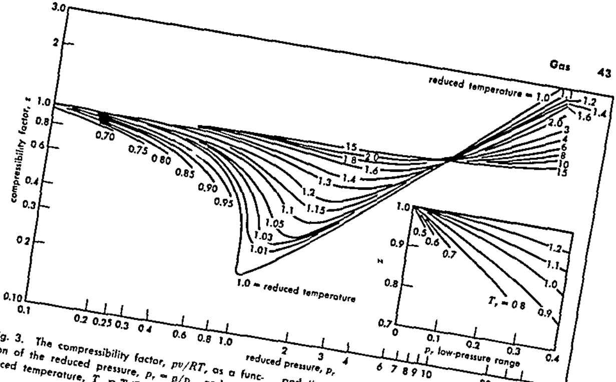


Fig. 3. The compressibility factor, pv/RT, as a function of the reduced pressure, $p_r = p/p_c$, and the reduced temperature, $T_r = T/T_c$. (From O. A. Hougen and K. M. Watson, Chemical Process Principles, part II, Wiley, 1947)

the critical point, z_c, should be a universal constant. It is found experimentally that this constant varies somewhat from one substance to another. On this account, A. L. Lydersen, R. A. Greenkorn, and O. A. Hougen have developed empirical tables of the compressibility factor and other thermodynamic properties of gases as functions of the reduced pressure and reduced temperature for a range of values of z_c. Such generalized corresponding-states treatments are very useful in predicting the behavior of a substance on the basis of scant experimental data.

Theoretical considerations. The equation-of-state behavior of a substance is closely related to the manner in which the constituent molecules interact. Through statistical mechanical considerations, it is possible to obtain some information about this relationship. If the molecules are spherically symmetrical, the force acting between a pair of molecules depends only on r, the distance between them. It is then convenient to describe this interaction by means of the intermolecular potential $\varphi(r)$ defined so that the force is the negative of the derivative of $\varphi(r)$ with respect to r. See INTERMOLECULAR FORCES.

Two theoretical approaches to the equation of state have been developed. In one of these approaches, the pressure is expressed in terms of the partition function Z and the total volume V of the container in the following manner:

$$p = kT(\partial \ln Z/\partial V)$$

Here k is the Boltzmann constant or the gas constant divided by Avogadro's number N_0, $k = R/N_0$. For a gas made up of spherical molecules or at with no internal structure, the partition function is

$$Z = \frac{1}{N!}\left(\frac{2\pi mkT}{h^2}\right)^{3N/2} \int \exp\left(-\sum_{i>j} \frac{\varphi_{ij}}{kT}\right) dv_1\, dv_2 \cdots dv_N$$

In this expression, φ_{ij} is the energy of interaction of molecules i and j and the summation is over all pairs of molecules, h is Planck's constant, N is the total number of molecules, and the integration is over the three Cartesian coordinates of each of the N molecules. The expression for the partition function may easily be generalized to include the effects of the structure of the molecules and the effects of quantum mechanics.

In another theoretical approach to the equation of state, the pressure may be written in the following manner:

$$p = \frac{NkT}{V} - \frac{2\pi N^2}{3V^2}\int g(r)\,\frac{d\varphi}{dr}\,r^3\,dr$$

In this expression $g(r)$ is the radial distribution function. This function is defined by the statement that $2\pi(N^2/V)g(r)r^2\,dr$ is the number of pairs of molecules in the gas for which the separation distance lies between r and $r + dr$. The radial distribution function may be determined experimentally by the scattering of x-rays. Theoretical expressions for $g(r)$ are being developed.

The compressibility factor $z = pV/NkT$ may be considered as a function of the temperature, T, and the molar volume, v. In the virial form of the equation of state, z is expressed as a series expansion in inverse powers of v,

$$z = 1 + B(T)/v + C(T)/v^2 + \cdots$$

The coefficients $B(T)$, $C(T)$, . . . , which are functions of the temperature, are referred to as the second, third, . . . , virial coefficients. This expansion is an important method of representing the deviations from ideal gas behavior. From statistical mechanics, the virial coefficients can be expressed in terms of the intermolecular potential. In particular, the second virial coefficient is

$$B(T) = 2\pi N_0 \int (1 - e^{-\varphi/kT}) r^2 \, dr$$

If the intermolecular potential is known, this expression provides a convenient method of predicting the first-order deviation of the gas from perfect gas behavior. The relation has often been used in the reverse manner to obtain information about the intermolecular potential. Often $\varphi(r)$ is expressed in the Lennard-Jones (6–12) form

$$\varphi(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Here ϵ and σ are constants characteristic of a particular substance. Values of these constants for many substances have been tabulated. In terms of these constants, the second virial coefficient has the form

$$B(T) = (2/3)\pi N_0 \sigma^3 B^*(kT/\epsilon)$$

where $B^*(kT/\epsilon)$ is a universal function. If all substances obeyed this Lennard-Jones (6–12) potential, the simple form of the law of corresponding states would be rigorously correct. *See* KINETIC THEORY OF MATTER; THERMODYNAMICS (CHEMICAL). [C.F.CU.; J.O.H.]

Bibliography: J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids*, 1954; A. L. Lydersen, R. A. Greenkorn, and O. A. Hougen, Univ. Wisconsin Eng. Exp. Sta. Rept. 4, Oct., 1955; O. A. Hougen and K. M. Watson, *Chemical Process Principles*, pt. 1, 2d ed., 1954, pts. 2 and 3, 1947.

Gas absorption operations

Processes in which a gas or gaseous component contained in an inert atmosphere is dissolved by a liquid. Occasionally, the gas to be absorbed is denoted as the absorbate, and the absorbing liquid as the absorbent.

Applications. Of the many thousands of gas absorption operations, only a few will be mentioned. For the production of inert gas atmospheres, synthesis gas, and ammonia, carbon dioxide must be removed from various accompanying components. On a commercial scale, carbon dioxide is usually removed by absorption in aqueous monoethanolamine solutions or hot potassium carbonate solution.

The removal of propane and butane from natural gas (scrubbing) is accomplished in plate towers, operating under several atmospheres pressure with an oil absorbent.

Removal of sulfur dioxide from flue gases and polluted atmospheres is an important operation. Because large gas quantities must be handled, grid-packed towers that operate under very low pressure drops and that have relatively small power requirements have been used with special alkaline solutions as absorbents.

As a final example of gas absorption, it may be mentioned that gas often is dried by contact with a desiccant solution. Thus, wet chlorine gas as obtained from electrolytic cells is passed through large packed-tower installations containing concentrated sulfuric acid. This reduces the water content in the effluent chlorine to a few parts per million.

Theory. In order to promote the absorption of the gas, a driving force is required. The gas concentration in the inert atmosphere is usually known. As the gas dissolves in the liquid, it will exert a certain back pressure. For a fixed temperature, this back pressure depends on the quantity of gas contained in the liquid, and it is known as the equilibrium pressure. The difference between the portion of the pressure (partial pressure) due to the gas in the atmosphere and the equilibrium pressure is a measure of the driving force. If the difference is positive, there will be absorption; if it is negative, desorption or stripping of the gas from the solution will prevail.

As in all transfer operations, resistances are encountered. These are found in both the gas and liquid parts of the system. The resistances are confined primarily to the contacting films of the phases at the interface. This simplified picture of the mechanism is used to provide a practical basis for gas absorption equipment design and for evaluation of mass transfer rates.

Finally, mass transfer requires a contacting surface between the liquid and gas phases. Obviously, larger and more accessible contacting surfaces will permit more rapid absorption. Equipment design is concerned largely, though not wholly, with this consideration.

Equipment. The most common absorption units are towers of either the plate or packed type, which permit countercurrent flow of liquid and gas streams.

Plate towers are of various designs. In so-called bubble-cap plates, the gas bubbles through a small head of liquid, and the liquid flows sideways across the plate and descends by way of downcomers to the plate below (Fig. 1). In the simpler sieve or perforated plates, gas-liquid passage is usually by way of small holes in the plates or between metal

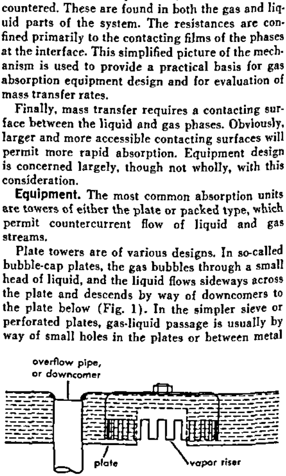


Fig. 1. Diagram of a plate and bubble cap from a plate tower.

Fig. 2. Random-packed Intalox saddles. (U.S. Stoneware Co.)

bars arranged in parallel. The construction of bubble-cap plates permits them to be operated over a wider range of liquid and gas rates than perforated plates and their various modifications.

A packed tower is essentially a vertical shell which carries a tower packing on a support plate and which embodies a liquid distributor above the packing. Tower packings are small individual shapes used to compose a packed bed (Fig. 2). It is the function of the tower packing to provide a skeleton along which the liquid passes in the form of a thin film and through the interstices of which the gas passes. Packed towers are of simpler construction than plate towers. However, in vessels of large diameter and with low prevailing liquid rates, a careful analysis frequently is required to determine the relative economies of the two.

Tower packings, manufactured chiefly of ceramics, metals, and plastics, may be classed as dumped and stacked types. The dumped packings include the various ring, saddle, and spiral shapes up to about 2-in. nominal size. Rings of 3-in. nominal size and larger are almost always stacked. Dumped beds are employed with relatively low liquid and gas rates where pressure drops must not be very small. Stacked beds, on the other hand, permit very high flow rates at quite low pressure drops. An interesting type of stacked tower is the grid-packed tower, of which certain water-cooling towers are examples.

Occasionally, spray towers are employed in gas absorption. Their capacity is limited by droplet size and prevailing gas velocity. Other gas-liquid contactors are comprised of tanks with impellers and similar constructions. Units of this type have not yet been standardized. *See* DISTILLATION; HUMIDIFICATION; MASS-TRANSFER OPERATION; STRIPPING. [M.L.]

Bibliography: T. K. Sherwood, and R. L. Pigford, *Absorption and Extraction*, 2d ed., 1952.

Gas analysis

Gases are analyzed by two general methods: absorptiometric methods, which involve chemical reactions with specific reagents; and instrumental methods, which utilize various physical properties of the sample components. The analysis of gases is important in process control, detection of leaks, flue and exhaust gas analyses, and research.

Absorptiometric methods. In classical absorptiometric analysis, a measured volume of sample at a specific temperature and pressure is exposed in a closed chamber to a suitable solid or liquid reagent which is capable of combining chemically, by mechanical solution, or by an adsorption phenomenon with one or more of the sample components. Removal of a constituent is generally determined either by returning the sample to its original pressure and measuring the loss in volume or by returning the sample to its original volume and measuring the decrease in pressure. The measuring vessels are usually water-jacketed to minimize errors due to temperature fluctuations during the analysis.

Individual absorption reagents will not remove one and only one constituent from any gas mixture, as represented by the ideal case, but rather will remove a certain type or class of gas. Thus, potassium hydroxide, a standard absorbent for carbon dioxide, will absorb other acidic gases, such as sulfur dioxide, as well. Therefore, careful attention must be given to the order in which the sample is exposed to the reagents. For complete analysis of complex mixtures, the sample must be broken down into fractions, for example, by fractional distillation, which are then analyzed separately.

For some gases for which there are no applicable specific absorbents, such as saturated hydrocarbons, an oxidation technique is sometimes employed. The relative volumes of gases originally present may then be determined from the amount of an oxidation product or from the loss in volume after the oxidation products are absorbed. In some cases, it is possible to determine two or more components by oxidizing them successively at different conditions. A combustion method proposed by Justus von Liebig in 1831 for determining the hydrogen and carbon content of organic material is still a standard procedure. The sample is burned by sweeping it with oxygen over hot copper oxide. The gaseous products are swept into a weighed tube of calcium chloride to determine the water formed and then into a weighed potassium hydroxide absorber to determine the carbon dioxide produced. There are only a few important gases for which oxidation methods, absorbents, or both are

The equipment for absorptiometric analysis may be classified according to whether the measurement is made volumetrically or manometrically, and according to the number and type of vessels used for the volume or pressure measurements, for the gas adsorption, and for the combustion. The so-called Orsat apparatus has separate vessels for the absorption and combustion which are permanently attached to the measuring vessel. In one modification of the Orsat apparatus, the reagents used are potassium hydroxide for acid gases, potassium iodomercurate for acetylene, maleic anhydride for dienes, acid mercuric sulfate for olefins, chromous chloride for oxygen, acid cuprous chloride for carbon monoxide, copper oxide wire at 270°C for hydrogen, and precipitated copper oxide containing 1% iron oxide at 700°C for combustion of saturated hydrocarbons.

Sulfuric acid in varied concentrations will differentiate between various types of olefinic hydrocarbons. Thus, 65% sulfuric acid separates isobutylene from other butylenes, 75% sulfuric acid separates tertiary amylene from other amylenes, and 87% sulfuric acid absorbs all olefinic hydrocarbons through the C_4's except ethylene.

When the amount of available sample is limited to 1 ml or less, an absorptiometric analysis must be made by low-pressure methods, using a vacuum apparatus or a capillary buret. Such specialized techniques are used for determining gases occluded by metals and glasses.

Routine absorptiometric equipment is not capable of making analyses below concentrations of about 0.05%. However, trace analysis techniques have been developed for determining components in concentrations as low as 0.01 part per million. These methods usually involve removal of the trace constituents from a large measured volume of the gas by a suitable adsorbent or by a freeze-out trap. The amounts of the separated constituents are then determined colorimetrically, titrimetrically, or gravimetrically. The principal adsorbents used in trace gas analysis are: for organic gases and water, activated charcoal, silica gel, fuller's earth, activated clays, bone char, and activated alumina; for water, anhydrous calcium chloride, calcium sulfate, magnesium and barium perchlorates, and phosphorus pentoxide; for carbon dioxide, soda lime and Ascarite; for hydrogen sulfide, asbestos impregnated with silver oxide; and for sulfur dioxide, porous solids in which manganese dioxide has been deposited.

Trace analyses are sometimes made conveniently with specific absorption reagents contained in a small tube through which the gas sample is forced by means of an aspirator bulb. The reagent reacts with the component of interest to produce a stain, the extent or depth of which permits an estimate of the amount of the component. In this way, car- [illegible] with a tube containing silver cyanide deposited on activated alumina, and hydrogen cyanide with a tube containing benzidine-copper acetate on an inert solid, both in the concentration range of a few parts per million.

Instrumental methods. These have been developed largely since about 1925 and are based upon the physical properties, rather than chemical properties, of gases. The instruments utilized range from simple resistance thermometers to highly complex spectrophotometers. Many instrumental methods are well-suited to the streamlining of plant operations so that operating variables can be controlled from a panel board. Physical properties that have been used as the basis for instrumental methods include freezing and boiling points, density, refractive index, electrical and thermal conductivity, magnetic susceptibility, effusion rate, sonic velocity, heat of combustion or reaction, molecular weight (mass spectrometry), and ultraviolet or infrared emission or absorption.

Low-temperature distillation, as developed by W. J. Podbielniak, is an elegant technique which has been widely used both for analysis and as a preparative step for other methods. The fractionating column has a vacuum jacket with a metal radiation shield and is cooled with liquid nitrogen. The gas mixture to be analyzed is passed into the bottom of the chilled evacuated column. The heater and the rate of flow of the liquid nitrogen coolant are then adjusted to provide suitable conditions for the distillation. The components are identified by boiling point (reflux temperature), and the amounts are determined at various points during the distillation by measuring the corresponding pressure increases in a previously evacuated receiver. The best apparatus is equipped with a special potentiometer-type strip chart recorder coupled with a manometer to plot directly the reflux temperature against pressure rise in the receiver, and it has a system of controlling all distillation variables, including the reflux ratio.

Since about 1953, gas chromatography has rapidly supplanted or supplemented many of the older methods. In this versatile new technique, the components are separated by passing the sample, usually by means of an inert carrier gas, through a column of a granular solid adsorbent or a supported liquid packing. The components emerge separately and are determined at the exit of the column, usually by thermal conductivity. Nearly all gas mixtures can be resolved and analyzed quantitatively by this technique. *See* GAS CHROMATOGRAPHY; MASS SPECTROSCOPE; SPECTROPHOTOMETRIC ANALYSIS; VACUUM FUSION. [F.T.E.]

Bibliography: V. J. Altieri, *Gas Analysis and Testing of Gaseous Materials*, 1945; P. W. Mullen, *Modern Gas Analysis*, 1955.

Gas chromatography

A separation technique involving passage of a gaseous moving phase through a column containing a fixed phase. It includes (1) gas-liquid chromatog-

raphy (GLC), or more precisely, gas-liquid partition chromatography, in which the fixed phase (column packing) is a liquid solvent distributed on an inert solid support; and (2) gas-solid chromatography (GSC), in which the fixed phase is a surface-active sorbent (charcoal, silica gel, activated alumina). Gas chromatography, which has come into prominence only since about 1953, is used principally as an analytical technique for the determination of volatile compounds (gases and liquids) with boiling points up to 400°C or even higher. However, it is useful also as a research method for determining certain physical quantities such as distribution or partition coefficients and adsorption isotherms, and as a preparative technique for isolating pure components or certain fractions from complex mixtures.

Gas chromatographic separations are achieved by selective retardation exerted by the fixed phase because of differences in partition coefficients (GLC) or in adsorption (GSC). As in other forms of chromatography, separations may be achieved, at least theoretically, by three different techniques: (1) elution development, (2) frontal analysis, and (3) displacement development. Elution development employs a small sample injected at the inlet

lium, nitrogen, or hydrogen. The band of each component travels through the column at a rate specific for that component under the conditions of the experiment. If the bands are separated, pure carrier gas emerges from the column between them. Displacement development uses a carrier gas which is saturated with a vapor of a strongly held substance, and all bands are moved forward by the displacer and issue from the column stepwise without intervening gaps. An essential condition for good separations by this technique is that the fronts of each band be steep and sharp. In frontal analysis, a relatively large sample is passed continuously into the column in gas phase, usually after admixture with a certain amount of an inert gas. The components emerge from the column according to their degree of retention by the column packing and without intervening gaps. Displacement development is scarcely feasible in GLC, and frontal analysis has no particular advantage. However, these techniques have both been used in GSC for special purposes. The information to follow will pertain to elution development, which is by far the simplest and most versatile technique.

Gas-liquid chromatography. GLC is more versatile than GSC because of the large number of available solvents having widely different separating characteristics, whereas in GSC there is relatively little choice of adsorbents. However, GSC is useful for separating very low-boiling components

of the adsorption isotherms. On the other hand, in GLC the distribution isotherms are usually linear, giving rise to symmetrical peaks of a Gaussian, or bell, shape.

The essential features of GLC are shown schematically in Fig. 1. The column consists of a glass or metal tube, which may be straight, U-shaped, or coiled, containing an inert, finely divided solid on which is supported the liquid stationary phase. The liquid phase must have a low volatility at the operating temperature. The column is maintained at a constant temperature in a liquid or air bath, or suitable furnace. The carrier gas is forced through the column at the desired (and constant) rate by a flow control device. When the conditions of operation have become stabilized, a sample of the mixture to be analyzed is injected, usually from a syringe, through a serum cap, into a vaporizing sec-

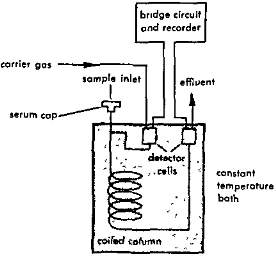


Fig 1. Schematic diagram of a gas chromatography apparatus.

tion at the front end of the column. The composition of the emerging gas (effluent) is monitored by a suitable detecting device capable of indicating the presence and amount of the individual components.

The two main functions of the process and equipment are (1) the separation of the constituents and (2) their detection and measurement. The separating power, or efficiency, of a column depends upon many factors, chief among which are the nature and amount of the stationary liquid, the type and particle size of the support, the length and diameter of the column, the temperature, gas velocity and pressure in the column, and the size and type of sample. For maximum separating power, these variables are adjusted so that during transport of the sample through the column the broadening of elution bands is held to a minimum. Band broadening is caused by longitudinal diffusion effects and resistance to mass transfer (nonequilibrium) between the gaseous and liquid phases. The width of the bands in relation to the retention time in the column provides a means of measuring

The equipment for absorptiometric analysis may be classified according to whether the measurement is made volumetrically or manometrically, and according to the number and type of vessels used for the volume or pressure measurements, for the gas adsorption, and for the combustion. The so-called Orsat apparatus has separate vessels for the absorption and combustion which are permanently attached to the measuring vessel. In one modification of the Orsat apparatus, the reagents used are potassium hydroxide for acid gases, potassium iodomercurate for acetylene, maleic anhydride for dienes, acid mercuric sulfate for olefins, chromous chloride for oxygen, acid cuprous chloride for carbon monoxide, copper oxide wire at 270°C for hydrogen, and precipitated copper oxide containing 1% iron oxide at 700°C for combustion of saturated hydrocarbons.

Sulfuric acid in varied concentrations will differentiate between various types of olefinic hydrocarbons. Thus, 65% sulfuric acid separates isobutylene from other butylenes, 75% sulfuric acid separates tertiary amylene from other amylenes, and 87% sulfuric acid absorbs all olefinic hydrocarbons through the C_4's except ethylene.

When the amount of available sample is limited to 1 ml or less, an absorptiometric analysis must be made by low-pressure methods, using a vacuum apparatus or a capillary buret. Such specialized techniques are used for determining gases occluded by metals and glasses.

Routine absorptiometric equipment is not capable of making analyses below concentrations of about 0.05%. However, trace analysis techniques have been developed for determining components in concentrations as low as 0.01 part per million. These methods usually involve removal of the trace constituents from a large measured volume of the gas by a suitable adsorbent or by a freeze-out trap. The amounts of the separated constituents are then determined colorimetrically, titrimetrically, or gravimetrically. The principal adsorbents used in trace gas analysis are: for organic gases and water, activated charcoal, silica gel, fuller's earth, activated clays, bone char, and activated alumina; for water, anhydrous calcium chloride, calcium sulfate, magnesium and barium perchlorates, and phosphorus pentoxide; for carbon dioxide, soda lime and Ascarite; for hydrogen sulfide, asbestos impregnated with silver oxide; and for sulfur dioxide, porous solids in which manganese dioxide has been deposited.

Trace analyses are sometimes made conveniently with specific absorption reagents contained in a small tube through which the gas sample is forced by means of an aspirator bulb. The reagent reacts with the component of interest to produce a stain, the extent or depth of which permits an estimate [illegible] with a tube containing silver cyanide deposited on activated alumina, and hydrogen cyanide with a tube containing benzidine-copper acetate on an inert solid, both in the concentration range of a few parts per million.

Instrumental methods. These have been developed largely since about 1925 and are based upon the physical properties, rather than chemical properties, of gases. The instruments utilized range from simple resistance thermometers to highly complex spectrophotometers. Many instrumental methods are well-suited to the streamlining of plant operations so that operating variables can be controlled from a panel board. Physical properties that have been used as the basis for instrumental methods include freezing and boiling points, density, refractive index, electrical and thermal conductivity, magnetic susceptibility, effusion rate, sonic velocity, heat of combustion or reaction, molecular weight (mass spectrometry), and ultraviolet or infrared emission or absorption.

Low-temperature distillation, as developed by W. J. Podbielniak, is an elegant technique which has been widely used both for analysis and as a preparative step for other methods. The fractionating column has a vacuum jacket with a metal radiation shield and is cooled with liquid nitrogen. The gas mixture to be analyzed is passed into the bottom of the chilled evacuated column. The heater and the rate of flow of the liquid nitrogen coolant are then adjusted to provide suitable conditions for the distillation. The components are identified by boiling point (reflux temperature), and the amounts are determined at various points during the distillation by measuring the corresponding pressure increases in a previously evacuated receiver. The best apparatus is equipped with a special potentiometer-type strip chart recorder coupled with a manometer to plot directly the reflux temperature against pressure rise in the receiver, and it has a system of controlling all distillation variables, including the reflux ratio.

Since about 1953, gas chromatography has rapidly supplanted or supplemented many of the older methods. In this versatile new technique, the components are separated by passing the sample, usually by means of an inert carrier gas, through a column of a granular solid adsorbent or a supported liquid packing. The components emerge separately and are determined at the exit of the column, usually by thermal conductivity. Nearly all gas mixtures can be resolved and analyzed quantitatively by this technique. *See* GAS CHROMATOGRAPHY; MASS SPECTROSCOPE; SPECTROPHOTOMETRIC ANALYSIS; VACUUM FUSION. [F.T.E.]

Bibliography: V. J. Altieri, *Gas Analysis and Testing of Gaseous Materials*, 1945; P. W. Mullen, *Modern Gas Analysis*, 1955.

Gas chromatography

A separation technique involving passage of a gaseous moving phase through a column containing a fixed phase. It includes (1) gas-liquid chromatog-

The chief advantages of gas chromatography over other methods of analysis are its great versatility and specificity, high separating power, and small sample requirement. The equipment is extremely simple and the analysis is relatively fast. *See* ADSORPTION; CHROMATOGRAPHY; EXTRACTION; GAS ANALYSIS. [F.T.E.]

Bibliography: A. I. M. Keulemans, *Gas Chromatography*, 1957; C. Phillips, *Gas Chromatography*, 1956.

Gas constant

Boyle's law and Charles' law may be combined into a single expression showing how the volume V of a given mass of gas depends upon its temperature T and pressure p, pV/T = a constant. If the mass of gas chosen is 1 gram-molecule, then the constant, known as the gas constant, is written as R. Hence, $pV = RT$ for 1 gram-mole. *See* BOYLE'S LAW; CHARLES' LAW; GAS.

The numerical value of the gas constant R is obtained by dividing the molar or gram-molecular volume of a perfect gas at the melting point of ice and a pressure of 76 cm of mercury at the same temperature (the standard atmosphere) by the absolute temperature at the ice-point, $R = V_0/T_0$. The best values available for R, in various units, are given below:

$$
\begin{aligned}
R &= 8.31436 \pm 0.00038 \times 10^7 \text{ erg/(deg)(mole)} \\
&= 1.986467 \text{ cal/(deg)(mole)} \\
&= 8.20544 \pm 0.00037 \times 10^{-2} \text{ liter-atm/(deg)(mole)}
\end{aligned}
$$

See MOLAR VOLUME.

The kinetic theory of gases relates R to the specific heats of a perfect monatomic gas at constant pressure C_p and constant volume C_v. C_p is equal to $\frac{5}{2}R$ and C_v to $\frac{3}{2}R$. Maxwell's law of the equipartition of energy shows that any degree of freedom of a system possesses an energy of $\frac{1}{2}RT$ per mole, and hence, makes a contribution of $\frac{1}{2}R$ to the specific heat. Thus, a perfect monatomic gas, which has three degrees of freedom, represented by the three independent directions of motion of its molecules, has a specific heat, C_v, of $\frac{3}{2}R$. A perfect, diatomic gas, on the other hand, whose molecules can rotate in two independent directions as well, has a specific heat C_v of $\frac{5}{2}R$. *See* KINETIC THEORY OF MATTER. [T.C.W.]

Gas dynamics

Dynamic motion of gases experiencing thermal effects. As a science, gas dynamics combines fluid mechanics and thermodynamics. It differs from gas statics in that there is motion and from gas kinematics in that forces exerted on or by the gas are taken into consideration.

Scope of subject. Several other names are used to define the subject. The most important ones are aerothermodynamics, aerothermochemistry, fluid dynamics, compressible fluid flow, and supersonic aerodynamics. The term aerothermodynamics has the shortcoming of limiting itself to air, although it specifically indicates that the thermodynamics of the fluid (air in this case) is to be considered. Aerothermochemistry, too, specifies air as the fluid and indicates that chemical effects are to be studied. However, this is not always so. Compressible fluid flow emphasizes that the fluid is compressible, in contrast to hydrodynamics; supersonic aerodynamics restricts itself to phenomena taking place in air at velocities faster than the speed of sound.

In the classical sense, gas dynamics excludes factors such as magnetism and electricity. However, in some reentry and propulsion applications, these effects must be considered. One then deals with magnetogas dynamics.

Classical gas dynamics is not concerned with the behavior of the individual particles (atoms or molecules, for example) making up the gas. Instead, it deals primarily with the motion of the gas as a mass or continuum. However, at high altitudes particles are far apart, and continuum considerations may not be applicable. To determine whether or not the gas is a continuum one must evaluate the Knudsen number Kn, which is defined as the ratio of the mean free path to the characteristic length of the body in the flow stream. Thus, if $Kn = 0.01$ the continuum gas dynamic equations may be applied. On the other hand, when $Kn = 1$ the flow is termed slip flow, and when $Kn = 10$ it is called free molecule flow. *See* SUPERAERODYNAMICS.

In gas-dynamic analyses it is customary to use the concept of the control volume. This is an arbitrarily defined volume which fluid may enter or leave (*see* FLUID MECHANICS). The fundamental laws of fluid mechanics and thermodynamics when applied to the control volume form the basis of gas dynamics.

Fundamental relations. In general, a problem in gas dynamics is solved if several equations can be satisfied—conservation equations, thermodynamic equations, and certain flow parameters. Nomenclature for these equations is defined in the list.

Nomenclature of gas dynamics

- A Cross-sectional area
- a Acoustic velocity
- C Constant, as in Eq. (28)
- c_p Specific heat at constant pressure
- c_v Specific heat at constant volume
- $\mathbf{F}_e$ Electric force
- $\mathbf{F}_g$ Gravity force
- $\mathbf{F}_m$ Magnetic force
- f Friction factor, as in Eq. (29)
- G Mass velocity, $G = \rho V$
- G' Mass flow per unit time
- h Enthalpy
- k Thermal conductivity
- L Characteristic length
- M Mach number, V/a
- n_i Number of particles of component i per unit volume
- p Pressure
- p_0 Total or stagnation pressure
- p^* Critical or throat pressure

p_o Original pressure
Q Heat
Q_j Joule heat
Q_R Radiant heat
q Dynamic pressure
R Gas constant, as in Eq. (12)
r_h Hydraulic radius
s Entropy
T Temperature
T_0 Total or stagnation temperature
T_o Original or initial temperature
t Time
u Internal energy
V Average gas velocity
$\mathbf{V}$ Gas vector velocity
$\mathbf{v}_i$ Particle vector velocity
w_i Rate of production of components i
x Per cent dissociation
α Mach angle
α Per cent ionization
γ Specific heat ratio, c_p/c_v
ν Kinematic viscosity
ρ Density
ϕ Work dissipated in losses
∇ Operator, del

Continuity equation. When relativistic considerations are nonexistent, mass cannot be created or destroyed. Two types of continuity equation may be written, namely the global and the species continuity equations. *See* CALCULUS OF VECTORS.

The global continuity equation is written

$$\nabla \cdot (\rho \mathbf{V}) + \frac{\partial \rho}{\partial t} = 0 \qquad (1)$$

The first term defines the changes with respect to the space coordinates (spatial differential), whereas the second indicates changes with respect to time, and thus, specifies the acceleration or deceleration. For steady flow (no acceleration) the continuity equation becomes

$$\nabla \cdot (\rho \mathbf{V}) = 0 \qquad (2)$$

whereas for incompressible flow it is

$$\nabla \cdot \mathbf{V} = 0 \qquad (3)$$

If the gas undergoes chemical changes, such as may occur in a combustion process inside a turbojet engine or behind a shock wave when a missile travels at extreme speeds, the conservation of the species must be accounted for. In this case the continuity equation is written as follows:

$$\frac{\partial n_i}{\partial t} + \nabla \cdot (n_i \mathbf{v}_i) = \dot{w}_i \qquad (4)$$

This equation must be written for each species and thus there will be i species continuity equations. The term on the right-hand side is the rate of production of the species. When Eq. (4) is summed for all species it will revert to the global form of the continuity equation.

Frequently a one-dimensional approach is followed in gas dynamics, in which case the properties are assumed to vary in the direction of flow only, or nearly so. The steady, one-dimensional, global continuity equation is written

$$\rho A V = \text{constant} = G' \qquad (5)$$

Momentum equation. The momentum equation expresses the conservation of momentum. It may take into consideration the effects of friction and external body forces such as gravity and magnetism. Other names for the momentum equation are the equation of motion, the Navier-Stokes equation, and Euler's equation. The momentum equation is generally written for the gross or global flow. The general form is given by

$$D\mathbf{V}/Dt = -\frac{1}{\rho}\nabla p + \nu \nabla^2 \mathbf{V} + (1/3)\nu \nabla (\nabla \cdot \mathbf{V}) + \mathbf{F}_g + \mathbf{F}_e + \mathbf{F}_m \qquad (6)$$

In this equation the term D/Dt represents the mobile operator which is defined as

$$D/Dt = (\partial/\partial t) + \mathbf{V} \cdot \nabla$$

The first term on the right-hand side represents the forces on the fluid control volume boundaries; the second and third terms on the right-hand side define the viscous effects; and $\mathbf{F}_g$, $\mathbf{F}_e$, and $\mathbf{F}_m$ represent, respectively, gravity forces, electric forces, and magnetic forces. Equation (6) is all-embracing and may be used in classical gas dynamics, in aerothermochemistry, and in magnetohydrodynamics. Written in vectorial form, Eq. (6) represents the various directions. Thus, in three-dimensional cartesian coordinates, the equation will have to be written for the x, y, and z directions. It may not be necessary to consider all terms. For example, if electromagnetic effects do not enter the case under study, the terms $\mathbf{F}_e$ and $\mathbf{F}_m$ are neglected (*see* NAVIER-STOKES EQUATION). Again, if inviscid flow is being considered, the second and third terms on the right-hand side are neglected (*see* EULER'S MOMENTUM THEOREM). Finally, if the flow under study is steady, the term $\partial/\partial t$ in the mobile operator is zero.

The momentum equation when written to include viscous effects is called the Navier-Stokes equation. No general solution exists for the Navier-Stokes equation, and thus each case must be solved individually. The momentum equation, when it is written for an inviscid fluid and when external forces such as magnetism and gravity are neglected, is called Euler's equation. Thus,

$$\frac{\partial \mathbf{V}}{\partial t} + (\nabla \times \mathbf{V}) \times \mathbf{V} + \nabla \frac{V^2}{2} = -\frac{1}{\rho}\nabla p \qquad (7)$$

In its one-dimensional form, Euler's equation for steady flow is

$$\frac{dp}{\rho} + V\,dV = 0 \qquad (8)$$

When Euler's equation is integrated one obtains Bernoulli's equation, which, because it is derived

from Euler's equation, is true for reversible processes. The incompressible Bernoulli equation is

$$\frac{p}{\rho}+\frac{V^2}{2}=\text{constant} \tag{9}$$

whereas for compressible isentropic flow one uses the equation

$$\frac{\gamma}{\gamma-1}\frac{p}{\rho}+\frac{V^2}{2}=\text{constant} \tag{10}$$

Energy equation. The energy equation defines the conservation of energy (*see* THERMODYNAMIC PRINCIPLES). In thermodynamics it is called the first fundamental law of thermodynamics. In its most general form, the energy equation is

$$\frac{dQ}{dt}=\frac{du}{dt}+\frac{d}{dt}(p/\rho)+\frac{d}{dt}\left(\frac{V^2}{2}\right)+Q_j+\frac{dQ_R}{dx}+\phi \tag{11}$$

In Eq. (11) the left-hand term defines the heat transferred across the control volume surfaces by conduction, namely the Fourier conduction. In turn the terms on the right-hand side denote respectively internal energy du/dt, the flow energy $d(p/\rho)/dt$, kinetic energy $d(V^2/2)/dt$, magnetic energy Q_j, the rate of energy flow by radiation dQ_R/dx, and the dissipation function ϕ which accounts for the work necessary in overcoming the losses.

As with the momentum equation, not all terms need to be considered in Eq. (11) if the physical problem does not include them. For example, if the flow is adiabatic the term dQ/dt would be zero. If electromagnetic effects are negligible, the Joule heat term Q_j will disappear. At low temperatures the radiation term will drop out. Finally, if no irreversibilities accompany the flow, then the dissipation term is zero.

Equation of state. The state of the gas in the control volume is determined by its thermal equation of state, which interrelates pressure p, density ρ, and temperature T. For a perfect gas the thermal equation of state is

$$p=\rho RT \tag{12}$$

At moderate pressure and temperature, air behaves as a perfect gas. In most of the next few sections it will be assumed that the gas is perfect. Deviations from a perfect gas will be considered later under real gas effects.

The properties of a fluid are defined by caloric equations of state. Thus the caloric equation of state for the isobaric (constant pressure) specific heat c_p may be written

$$c_p=c_p(T,p) \tag{13}$$

Dimensionless parameters. *In gas dynamics,* dimensionless similarity parameters are widely used. One of these (Knudsen number) has been mentioned.

The dimensionless parameters are convenient because they may be used in comparing wide ranges of experimental data, in predicting performance of a model under study and in simplifying the solution of analytical problems by similarity considerations. *See* DIMENSIONAL ANALYSIS; DYNAMIC SIMILARITY.

Acoustic velocity. A particularly useful dimensionless parameter is Mach number M, defined as the ratio of local flow velocity V to local acoustic velocity a. Thus

$$M=V/a \tag{14}$$

Gas-dynamic flow regimes can then be classified as

$M<1$	subsonic flow
$M=1$	sonic flow
$0.9<M<1.1$	transonic flow
$M>1$	supersonic flow
$M>5$	hypersonic flow

The acoustic velocity a is the velocity with which a small disturbance or pressure wave is propagated in a medium. Thus the velocity of sound and the Mach number are related to the compressibility of the medium. The acoustic velocity in a gas is given by

$$a=\sqrt{\left(\frac{\partial p}{\partial\rho}\right)_{s=\text{constant}}} \tag{15}$$

where the partial differential of p with respect to ρ at constant entropy s indicates that the disturbances must be small. For a perfect gas the acoustic velocity can be calculated from the equation

$$a=\sqrt{\gamma RT} \tag{16}$$

Consider a point source traveling with a velocity $V<a$ as shown in Fig. 1. At reference time zero the point disturbance is at A. At unit time later $1t$, the point source will have moved to B, a distance Vt. At two time units later $2t$, the source will have moved to point C, a distance $2Vt$. Meanwhile, the spherical-pressure wavefront which started from A will have grown after time $2t$ to have a radius $2at$, and the wavefront from B, which could not start until the source reached B, will have grown to

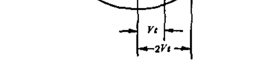


Fig. 1. Wavefronts produced by a point source moving at subsonic velocity. (A. B. Cambel and B. H. Jennings, Gas Dynamics, McGraw-Hill, 1958)

a radius at. At time $2t$ the source will have just reached point C. This step-by-step analysis of a point source moving at subsonic velocity indicates that each succeeding wavefront will always be contained within the preceding wavefront sphere.

If the point source moves so fast that $V > a$, the wavefront circles will no longer contain the source. The condition is shown in Fig. 2. The envelope to this family of circles is a straight line, known as the Mach line, and the Mach angle is such that $\sin \alpha = a/V$.

The Mach line constitutes a demarcation. In Fig. 2 the fluid outside the Mach line will receive no signal from the source. T. von Kármán has appropriately called this phenomenon the rule of forbidden signals and designated the region ahead of the Mach lines the zone of silence and the region inside the Mach lines the zone of action. *See* HYPERSONIC FLIGHT; SUBSONIC FLIGHT; TRANSONIC FLIGHT.

Stagnation temperature. If a thermometer is introduced into a nozzle through which a gas flows at high velocity, the thermometer reading will not be the true gas (free-stream or static) temperature. This is so because the gas immediately surrounding the thermometer sensing element is brought to rest. At the narrow region where the gas is brought completely to rest, the local temperature is the stagnation or total temperature. In this region the kinetic energy of the gas is converted into thermal energy. To explain this phenomenon, consider that a gas ejected from a reservoir at temperature T_o passes a thermometer at V_t and T_t. The energy equation is integrated as follows:

$$\frac{\gamma R}{\gamma - 1} \int_{T_o}^{T_t} dT + \int_{V_o=0}^{V_t=0} dV = 0 \qquad (17)$$

Therefore, $T_t = T_o$.

Because for a perfect gas enthalpy h is $dh = c_p\, dT$, theoretically the enthalpy of the gas brought to rest at the thermometer equals the enthalpy in the reservoir; therefore this is called the stagnation enthalpy. Furthermore, the temperature read by the stationary thermometer is theoretically equal to the temperature read in the reservoir. Irrespective of the location of the thermometer in this system, the stagnation conditions are set by the reservoir conditions. For a thermometer to indicate the static gas temperature it is necessary that it travel at the same velocity as the gas. This is impractical and hence the static temperature of the gas is generally determined from indirect measurements.

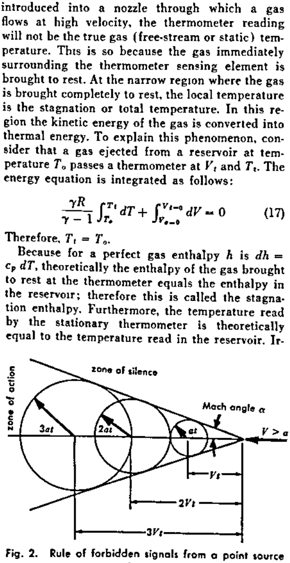


Fig. 2. Rule of forbidden signals from a point source moving at supersonic velocity. (*A. B. Cambel and B. H. Jennings, Gas Dynamics, McGraw-Hill, 1958*)

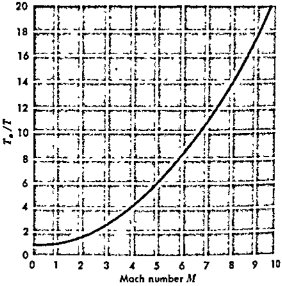


Fig. 3. Stagnation temperature rise ratio, for a perfect diatomic gas $\gamma = 1.4$. (*A. B. Cambel and B. H. Jennings, Gas Dynamics, McGraw-Hill, 1958*)

Only the gas at the stagnation point on the thermometer is brought to rest. The remainder travels at high velocity and the energy equation is then written as

$$c_pT + V^2/2 = c_pT_0 \qquad (18)$$

Therefore, the stagnation temperature is given by

$$T_0 = T + \frac{V^2}{2c_p} \qquad (19)$$

Because it is practical to determine the Mach number experimentally, the stagnation temperature may be written as a function of the Mach number. Thus dividing Eq. (19) by the acoustic velocity throughout and rearranging, one obtains the following important equation

$$\frac{T_0}{T} = 1 + \frac{\gamma - 1}{2} M^2 \qquad (20)$$

This equation is shown graphically in Fig. 3. It should be noted that unless specifically designed for the purpose, a stagnation thermometer will read a slightly different temperature from the theoretical stagnation temperature because of boundary-layer effects and heat losses. Thus each thermometer is ascribed a recovery factor which corrects for this deviation. Finally, at extreme Mach numbers ($M \geqq 5$) it is necessary to use a specific heat ratio γ other than the perfect gas value to take care of the associated real gas effects.

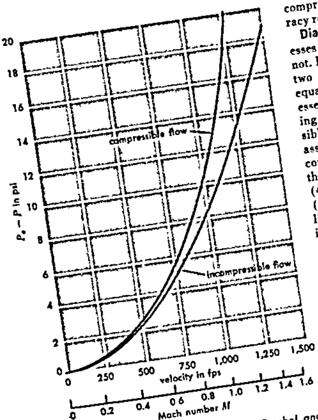


Fig. 4 Stagnation pressure rise. (A. B. Cambel and B. H. Jennings, *Gas Dynamics*, McGraw-Hill, 1958)

Total pressure. A relation between stagnation pressure and static pressure may be had by using certain of the perfect gas relationships. For perfect gas undergoing isentropic flow, the relationship between pressure and temperature is

$$\left(\frac{p_0}{p}\right)^{(\gamma-1)/\gamma} = \frac{T_0}{T}$$

Hence
$$\frac{p_0}{p} = \left(1 + \frac{\gamma - 1}{2} M^2\right)^{\gamma/(\gamma-1)} \qquad (21)$$

The dynamic pressure q is defined as

$$q = \rho V^2/2 = \gamma p M^2/2 \qquad (22)$$

Introducing the dynamic pressure into Eq. (21) and expanding in series form gives, for compressible flow

$$p_0 = p + \frac{V^2}{2}\left[1 + \frac{M^2}{4} + \frac{(2-\gamma)M^4}{24} + \cdots\right] \qquad (23)$$

In contrast, for incompressible flow, Bernoulli's equation for the stagnation condition is

$$p_0 = p + V^2/2 \qquad (24)$$

These two equations are compared graphically in Fig. 4.

One of the important cases for which the phenomena of Fig. 4 are applicable is the aircraft-speed indicator (*see* PITOT TUBE). If the speed indicator is calibrated without taking the effect of compressibility into consideration, serious inaccuracy results.

Diabatic flow. Although many engineering processes can be assumed to be adiabatic, others cannot. Heat exchangers and combustion chambers are two devices in which heat transfer occurs. The equations describing nonadiabatic or diabatic processes are complicated; consequently certain limiting assumptions are usually required to make possible analytical solutions of the equations. These assumptions are that (1) the flow takes place in a constant-area section, (2) there is no friction, (3) the gas is perfect and has constant specific heats, (4) the composition of the gas does not change, (5) there are no devices in the system which deliver or receive mechanical work, and (6) the flow is steady.

These assumptions are somewhat restrictive, but they yield results which, within a small margin of error, agree with actual values. Equations which conform to these requirements are called Rayleigh equations, and the associated flow is designated as Rayleigh flow. Figure 5 shows a model of Rayleigh flow. The energy equation for Fig. 5 may be written

$$\begin{aligned} Q_{1\to 2} &= c_p(T_2 - T_1) + (V_2^2 - V_1^2)/2 \\ &= h_2 - h_1 + (V_2^2 - V_1^2)/2 \end{aligned} \qquad (25)$$

If the stagnation enthalpy is introduced,

$$Q_{1\to 2} = h_{02} - h_{01} \qquad (26)$$

This in turn can be expressed in terms of stagnation temperatures

$$Q_{1\to 2} = c_p(T_{02} - T_{01}) \qquad (27)$$

Because for diabatic flow $Q_{1\to 2} \neq 0$ and because $c_p > 0$ always, it follows that $T_{02} \neq T_{01}$. This inequality states that in diabatic flow the stagnation

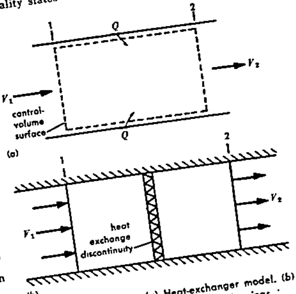


Fig. 5. Rayleigh flow. (a) Heat-exchanger model. (b) Heat addition. (A. B. Cambel and B. H. Jennings, *Gas Dynamics*, McGraw-Hill, 1958)

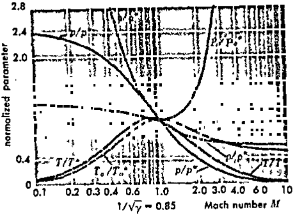


Fig. 6. Diabatic flow parameters for specific heat ratio of 1.4. Asterisk refers to condition where $M = 1$. (*A. B. Cambel and B. H. Jennings, Gas Dynamics, McGraw-Hill, 1958*)

temperature is not solely determined by the reservoir conditions, as is the case with adiabatic flow. Heating raises the stagnation temperature, and cooling lowers it.

The locus of points of properties during a constant-area, frictionless flow with heat exchange is called the Rayleigh line. By definition along the Rayleigh line the continuity equation and the momentum equation must apply. Thus, Eq. (5) applies to steady flow in a constant-area duct. Mass velocity by definition is $G = \rho V$, and from Eq. (6), the momentum relation is $p + \rho V^2 = C$, where C is a constant. Consequently

$$p + \frac{G^2}{\rho} = C \tag{28}$$

which is one of the many Rayleigh-line equations.

The variations of pressure, temperature, and density with Mach number for Rayleigh flow are plotted in Fig. 6. The fact that the curve for T_0/T_0^* in Fig. 6 reaches a maximum at a Mach number of unity indicates that it is impossible to pass from one flow domain into the other by the same heat-transfer process. Thus, if heat is added to a subsonic flow, the flow can be accelerated only until its Mach number becomes unity. Further addition of heat will not further accelerate the gas but will result in choking of the flow. As a consequence, the flow must readjust itself, which it will do by lowering its initial Mach number. Table 1 summarizes some of the Rayleigh flow phenomena.

Flow with friction. In practical devices the flow is accompanied by friction resulting in a pressure drop. Over a length dx, this pressure drop dp is given by the Fanning equation

$$dp = -4f\frac{\rho V^2}{2r_h}\,dx \tag{29}$$

where f is a friction factor that must be determined experimentally. To solve friction-flow problems analytically, certain simplifying assumptions are made and the resulting hypothetical flow is called Fanno flow. The Fanno flow assumptions are that (1) the cross-sectional area of the duct confining the gas is constant throughout (however, the section does not need to be circular as long as some characteristic diameter or hydraulic radius r_h can be defined for it); (2) the flow is adiabatic; (3) the gas is perfect and has constant specific heats; (4) the flow is steady; and (5) there are no devices in the system which deliver or receive mechanical work. Numerous Fanno flow equations may be written by combining the energy and the continuity equations in accordance with the above assumptions. In Fig. 7 may be seen the variation of properties during Fanno flow. Table 2 summarizes the trends of the most important properties during subsonic and supersonic flow.

When, for the same constant mass velocity $G = \rho V$, the Fanno and Rayleigh lines are plotted, the

Table 1. Variation of flow properties for Rayleigh flow

Property	Heating		Cooling	
	$M > 1$	$M < 1$	$M > 1$	$M < 1$
T_0	Increases	Increases	Decreases	Decreases
p	Increases	Decreases	Decreases	Increases
p_0	Decreases	Decreases	Increases	Increases
V	Decreases	Increases	Increases	Decreases
T	Increases	Increases when $M < 1/\sqrt{\gamma}$ Decreases when $M > 1/\sqrt{\gamma}$	Decreases	Decreases when $M < 1/\sqrt{\gamma}$ Increases when $M > 1/\sqrt{\gamma}$

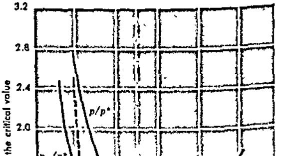


Fig. 7. Functions for constant-area flow with friction ($k = 1.4$). Asterisk refers to condition where $M = 1$. (*A. B. Cambel and B. H. Jennings, Gas Dynamics, McGraw-Hill, 1958*)

Table 2. Fanno flow phenomena

Property	Initial flow is subsonic	Initial flow is supersonic
M	Increases	Decreases
V	Increases	Decreases
p	Decreases	Increases
T	Decreases	Increases
ρ	Decreases	Increases

curves will appear as in Fig. 8. The Rayleigh and Fanno lines have two points of intersection denoted by a and b. A normal shock wave connects these two points. Because a shock wave constitutes an irreversibility, there will be associated with it an increase in entropy. Thus, point b will always lie to the right of point a.

There is another interesting point about the Fanno curve. If frictional flow continues along the subsonic portion of the Fanno line, the Mach number will tend to increase towards unity, whereas if it continues along the supersonic portion the Mach number will decrease towards unity. As in the case of Rayleigh flow, it is impossible, by virtue of the second law of thermodynamics, to pass from one flow regime to the other (subsonic into supersonic or vice versa) unless the mass velocity is *readjusted*.

Wave phenomena. When there is relative motion between a body and a fluid, the disturbance caused by the body (if infinitely small) is propagated through the fluid with the speed of sound. The speed of sound a in such cases is the speed with which rarefactions or compressions of infinitesimal amplitude propagate.

Shock wave. When the compressions in the flow are of finite amplitude, there usually occurs a discontinuous rise of pressure, and a shock wave exists. The propagation velocity a' of a shock wave will be higher than that of a sound wave. Figure 9 shows the ratio of the propagation velocity a' to the sound velocity a, with increases in pressure ratio across a discontinuity where p_1 and p_2 represent,

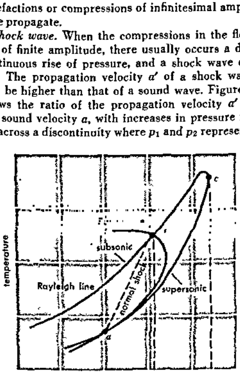


Fig. 8. Rayleigh and Fanno lines. (A. B. Cambel and B. H. Jennings, *Gas Dynamics*, McGraw-Hill, 1958)

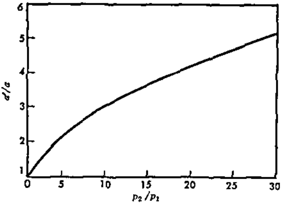


Fig. 9. Disturbance velocity ratio for air. (A. B. Cambel and B. H. Jennings, *Gas Dynamics*, McGraw-Hill, 1958)

respectively, the pressures upstream and downstream of the disturbance (*see* SHOCK WAVE). If initially the air is still and the shock wave is moving, after the passing of the shock the air will move in the direction of the shock. Gas compressions which have a finite amplitude travel faster than the speed of sound, as is the case in intense explosions (*see* WAVE MOTION IN FLUIDS). Generally variations in flow density are observed by optical laboratory methods (*see* SHOCK-WAVE DISPLAY). However, pilots of high-speed aircraft report seeing shock waves springing out from the wings of their aircraft.

Although shock waves are wave phenomena, it should be emphasized that not all wave phenomena are shock occurrences. Wave phenomena can be arbitrarily and qualitatively classified according to strength as weak waves and shock waves. A Mach wave, for example, is a weak wave. There are weak compression waves and weak expansion waves. The former may occur at a wall having a concave corner. In contrast, a weak expansion wave may occur at a convex wall corner. With very weak waves, it may be assumed that there are no drastic discontinuities because the deflection angle is infinitesimally small, and isentropic flow analysis is applicable across such a wave. Across a compression wave the flow is decelerated and the gas density is increased, and across an expansion wave the flow is accelerated. A shock wave, on the other hand, can only be of the compression type, and there is no rarefaction or expansion shock because this would require that the entropy decrease in an adiabatic process. Waves exhibiting entropy increases are classed as shock waves, whereas waves across which the entropy is substantially constant are referred to simply as waves.

Another manner of classifying waves, according to their relative velocity, differentiates them as stationary shock waves in contrast to nonstationary shock waves. A shock wave may be rendered stationary if the mass of gas in which it travels at a speed equal to that of the shock opposite in direction. Such occurrences may

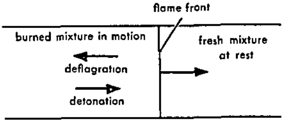


Fig. 10. Model for detonation and deflagration. (A. B. Cambel and B. H. Jennings, Gas Dynamics, McGraw-Hill, 1958)

served in nozzles, for example. Stationary shock waves may be either attached to an object or detached, depending upon the shape of the object and the Mach number. Nonstationary shock waves can readily be developed in shock tubes.

Finally, shock waves may be situated at right angles to the flow, in which case they are called normal shocks, or they may be located at an angle, in which case they are called oblique shocks. Normal shocks may be treated by a one-dimensional flow analysis, but oblique wave phenomena (whether shock waves or weak waves) require two-dimensional flow analysis.

Across a shock wave the fundamental flow equations of conservation of mass, momentum, and energy apply. For example, across a normal shock, the velocity and the total pressure decrease while the pressure, the temperature, and density increase. However, the stagnation temperature is the same on each side of a normal shock. Finally, although upstream of the normal shock the flow is supersonic, it is always subsonic downstream of the shock. *See* Diffuser.

Combustion wave. One peculiar type of wave phenomenon is that associated with a chemical reaction. Assume that a combustible mixture is contained in a tube and is subsequently ignited so that a flame front propagates downstream (*see* Combustion wave measurement). A model of this phenomenon is depicted in Fig. 10. Whether detonation or deflagration occurs depends upon the conditions summarized in Table 3. Deflagration and detonation phenomena are generally studied by the Chapman-Jouget analysis of Fig. 11. According to classical Chapman-Jouget analysis a weak detonation and a strong deflagration are excluded. In Fig. 11 there are two curves, namely one for $Q = 0$ and the other for $Q > 0$. The first is a wave phenomenon represented by

$$u_2 - u_1 = \frac{(p_1 + p_2)(\rho_2 - \rho_1)}{2\rho_1\rho_2} \tag{30}$$

whereas the second one is represented by the Hugoniot equation

$$u_b - u_u - Q = \frac{(p_u + p_b)(\rho_b - \rho_u)}{2\rho_u\rho_b} \tag{31}$$

where subscripts are as defined in Table 3.

Table 3. Detonation and deflagration*

	M_u	M_b	p_b/p_u	V_b/V_u	ρ_b/ρ_u	T_b/T_u
Detonation	>1	≧1	>1	<1	>1	>1
Deflagration	<1	>1	<1	>1	<1	>1

* Subscripts u and b refer respectively to unburned and burned mixtures

Variable area flow. In engineering devices it is often desired to accelerate or decelerate the flow. If the flow is accelerated the configuration is termed a nozzle. In most engineering applications the contour of the nozzle is first converging and then diverging. The narrowest area of such a converging-diverging or de Laval nozzle is called the throat and the exit is called the mouth (*see* Nozzle). For a convergent-divergent nozzle in which the flow is isentropic, the velocity at the throat is sonic, or in other words, the Mach number equals unity. The throat pressure is then said to be critical p_c and is given by

$$p_c = p_o\left(\frac{2}{\gamma + 1}\right)^{\gamma/(\gamma-1)} \tag{32}$$

The velocity at any point where the pressure is p within an idealized nozzle is given by

$$V = \frac{2\gamma R}{\gamma - 1} T_o[1 - (p/p_o)^{(\gamma-1)/\gamma}] \tag{33}$$

Consider a convergent-divergent nozzle inserted between two reservoirs as in Fig. 12. There will be no flow if the ratio of exit pressure p_e to reservoir pressure p_0 is $p_e/p_0 = 1$, case a in Fig. 13. If p_e is reduced so that it is slightly less than the entrance pressure, the convergent-divergent nozzle will act like a conventional venturi tube as represented by curve b in Fig. 13. For this case, the flow is always subsonic and resembles incompressible flow. When the exit pressure is reduced further, the critical pressure can be reached at the throat as curve g shows. In this case, the velocity is sonic at the throat, but is never supersonic within the

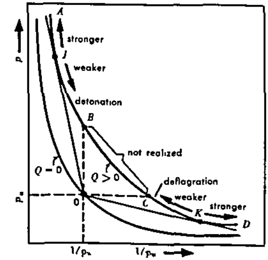


Fig 11. Chapman-Jouget conditions. (A. B. Cambel and B. H. Jennings, Gas Dynamics, McGraw-Hill, 1958)

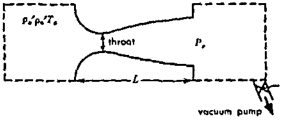


Fig. 12. Convergent-divergent nozzle. (A. B. Cambel and B. H. Jennings, *Gas Dynamics*, McGraw-Hill, 1958)

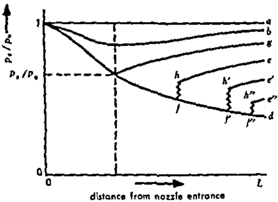


Fig 13. Pressure distribution in convergent-divergent nozzle. (A. B. Cambel and B. H. Jennings, *Gas Dynamics*, McGraw-Hill, 1958)

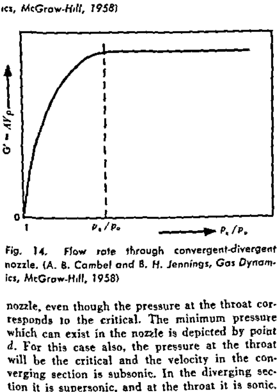


Fig. 14. Flow rate through convergent-divergent nozzle. (A. B. Cambel and B. H. Jennings, *Gas Dynamics*, McGraw-Hill, 1958)

nozzle, even though the pressure at the throat corresponds to the critical. The minimum pressure which can exist in the nozzle is depicted by point d. For this case also, the pressure at the throat will be the critical and the velocity in the converging section is subsonic. In the diverging section it is supersonic, and at the throat it is sonic. For the range of exit pressures from p_d to p_g the rate of flow curve is the same and is plotted in Fig. 14. The flow rate reaches a maximum value and remains at this value over this wide range of exit pressures.

The convergent-divergent nozzle in the absence of friction can give truly isentropic performance for the range of exhaust pressures from p_a to p_g and at the pressure reached along curve d, but not at intermediate pressures. The pressures p_g and p_d are the significant mouth-design pressures for a given nozzle. For the exhaust pressure range between p_d and p_g, flow discontinuities (shock) must occur in the nozzle such as show from f to h, followed by diffusion (pressure rise) from and after the shock point to an exit pressure such as p_e.

Real gas effects. In a perfect gas the molecules do not attract or repel one another. In addition, the *size of the molecules can be neglected and the gas* properties should not vary by other than temperature. If the gas under consideration does not obey these assumptions or if mixtures are being considered, then the perfect gas equation of state, Eq. (12), must be corrected by the introduction of a compressibility or departure coefficient Z (*see* VIRIAL EQUATION). Thus Eq. (12) becomes

$$p = \rho ZRT$$

The value of Z depends upon the gas and its condition.

If a gas is accelerated to high velocities, as may occur around hypersonic vehicles, it may dissociate into its neutral atoms. In this case the equation of state Eq. (12) becomes

$$p = \rho RT(1 + x)$$

where x is the fraction of dissociated gas.

If a gas is heated to sufficiently high temperatures, then its neutral atoms may be stripped of their electrons, and the gas becomes ionized. An ionized gas which has the same number of electrons and positive ions so that the net charge is neutral is a plasma (*see* PLASMA PHYSICS). For a plasma the equation of state becomes

$$p = \rho RT(1 + \alpha)$$

where α is the fraction of ionized gas.

In most flow conditions, the air surrounding the object is not homogeneous but a mixture of molecules, atoms, ions, and electrons (*see* MAGNETOGAS DYNAMICS). For a complex consisting of dissociated gas *and plasma the equation of state becomes*

$$p = \rho RT(1 + x + \alpha)$$

Great caution must be exercised in analyzing hypersonic flow because real gas effects (vibrational states, dissociation, electronic excitation, and ionization) will alter the flow greatly. *See* MAGNETOHYDRODYNAMICS.

In studying real gas phenomena one must determine whether the flow is thermodynamically in equilibrium or is frozen. This is done by evaluating the Damköhler number or Damköhler's first ratio which is defined

$$Dam = \frac{t_{transit}}{t_{relaxation}}$$

where $t_{transit}$ is the time the gas is in transit and $t_{relaxation}$ is the time required for a gas to return to its equilibrium state. If $Dam < 1$, frozen flow applies, whereas if $Dam > 1$, equilibrium flow applies.

The chemistry of real gases is complex because a correct explanation of it requires the inclusion of the behavior and characteristics of gases on the particle scale. *See* STATISTICAL MECHANICS; *see also* FLUID-FLOW PRINCIPLES; FLUID-FLOW PROPERTIES; HYDRODYNAMICS. [A.B.C.]

Bibliography: A. B. Cambel and B. H. Jennings, *Gas Dynamics*, 1958; H. W. Liepmann and A. Roshko, *Elements of Gasdynamics*, 1957; A. H. Shapiro, *The Dynamics and Thermodynamics of Compressible Fluid Flow*, 2 vols., 1953 and 1954.

Gas field and gas well

Petroleum gas, one form of naturally occurring hydrocarbons of petroleum, is produced from wells that penetrate subterranean petroleum reservoirs of several kinds. Oil and gas production are commonly intimately related, and about one-third of gross gas production is reported as derived from wells classed as oil wells. If gas is produced without oil, production is generally simplified, in part at least because the gas flows naturally without lifting, and also because of fewer complications in reservoir problems. As for all petroleum hydrocarbons, the term field designates an area underlain with little interruption by one or more reservoirs of commercially valuable gas.

See MINERAL FUEL AREAS; NATURAL GAS; OIL AND GAS FIELD DEVELOPMENT; OIL AND GAS FIELD EXPLOITATION; OIL AND GAS WELLS; PETROLEUM; PETROLEUM GEOLOGY; PETROLEUM RESERVOIR ENGINEERING.

Gas furnace

An enclosure in which a gaseous fuel is burned. Domestic heating systems may have gas furnaces. Some industrial power plants are fired with gases that remain as a by-product of other plant processes. Utility power stations may use gas as an alternate fuel to oil or coal, depending on relative cost and availability. *See* FURNACE (STEAM GENERATING). Some heating processes are carried out in gas-fired furnaces.

Among the gaseous fuels are natural gas, producer gas from coal, blast furnace gas, and liquefied petroleum gases such as propane and butane. Crude industrial heating gases carry impurities that corrode or clog pipes and burners. Solid or liquid suspensoids are removed by cyclones or electrostatic precipitators; gaseous impurities are removed chemically. The cleaned gas may be mixed with air in the furnace, in the burners, or in a blower before going to the burners. The gas and air may be supplied at moderate pressure, or one or both at high pressure. The high pressure component may serve to induce the other component into the furnace. The burner may be a single center-fire type or a multispud type with numerous small gas parts depending on how the heat is to be concentrated or distributed in the furnace. Crude uncleaned gases are fired through burners with large ports, the burners being removable for cleaning. *See* FUEL GAS. [R.M.H.]

Gas kinematics

The motion of a gas considered by itself, without regard for the cause of the motion. Various flow phenomena arise in widely different applications, yet the phenomena dealt with independently of their causes can be classed into relatively few fundamental patterns. Gas kinematics then deals with these abstracted flows. *See* GAS DYNAMICS; SINK FLOW; SOURCE FLOW. [F.H.R.]

Gas kinetics

The effects of gases due to motion. The simplest gas phenomena are those associated with gas at rest (*see* AEROSTAT; BUOYANCY; FLUID STATICS). Additional phenomena contribute to the effects when the gas is in motion (*see* AERODYNAMIC FORCE; BERNOULLI'S THEOREM; EULER'S MOMENTUM THEOREM; VISCOUS FLOW). The method of science is to isolate one phenomenon at a time for study, holding all else constant. However, in actual gas flows, motion is accompanied by other phenomena so that, for practical purpose, a more comprehensive view is necessary. *See* COMPRESSIBLE FLOW; GAS DYNAMICS; SUPERSONIC FLIGHT; TRANSONIC FLIGHT. [F.H.R.]

Gas mask

A device to protect the eyes and respiratory tract from noxious gases, vapors, and aerosols. Before World War I, the gas mask was unknown. The advent of chemical warfare necessitated the development of this protective device, which underwent an amazing transformation from a simple cotton cloth pad to an efficient military gas mask within 3½ years.

Three general types of gas masks are in common use, depending on the principle of operation utilized. The military protective mask and the commercial gas mask furnish clean air to the wearer by removing the contamination from the ambient atmosphere by means of a filter and a bed of adsorbent material. Rebreathing masks use closed systems which absorb the carbon dioxide and water from exhaled air and generate oxygen for inhalation. Air-supply masks furnish clean air from a source independent of the surrounding atmosphere.

Protective masks. The military protective mask is made up of three essential components—the facepiece, the canister, and the carrier. In the commercial gas mask the carrier is sometimes not used.

The facepiece covers the face, and sometimes the whole head. It consists of a rubber faceblank containing plastic or glass eyepieces for vision, an opening for incoming air, an outlet valve for exhaled air, and a nosecup for separating the incoming from the exhaled air. Purified air enters the mask through the canister, which may be attached to the facepiece either directly or by means of a

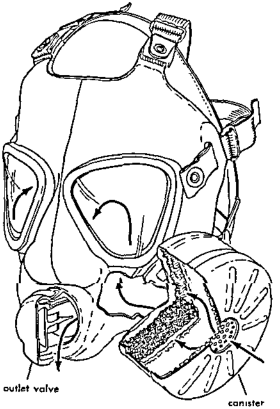


Fig. 1. The protective mask. Arrows indicate direction of air flow

flexible tube. In the United States military protective mask and in some commercial masks of advanced design, the air then passes over the inside of the eyepieces before it is inhaled. Exhaled air is exhausted through the outlet valve. The passage of air over the eyepieces aids in preventing fogging. The nosecup further assists in this matter by keeping the saturated exhalations from the lenses. The facepiece is kept in place by a self-centering head harness.

Canister adsorbents may be either activated charcoal or a mixture of charcoal and reactive chemicals. The charcoal is made in the form of hard granules to withstand rough usage and, in the United States military mask, is impregnated with several metal salts. These metals, principally silver, copper, and chromium, are necessary as decomposition catalysts. The toxic vapors of high-molecular-weight substances are readily adsorbed on activated charcoal and are not easily desorbed. This is not true of gases of low molecular weight, which must be decomposed to be rendered harmless. The metal salt impregnation performs this service Special-purpose commercial canisters sometimes contain reactive chemicals such as soda lime or sodium carbonate in addition to the charcoal, to aid in removing toxic gases.

Presently known activated or impregnated charcoals are extremely efficient in adsorbing practically all toxic vapors. A notable exception is carbon monoxide. To eliminate carbon monoxide the catalytic action of hopcalite is used to convert it to carbon dioxide by reaction with atmospheric oxygen. Hopcalite is a mixture of 50% manganese dioxide, 30% copper oxide, 15% cobaltic oxide, and 5% silver oxide. In masks to be worn in carbon monoxide atmospheres, a layer of hopcalite is placed in the canister below the adsorbent bed. Since hopcalite is adversely affected by moisture, a layer of a granular desiccant is placed between the hopcalite and adsorbent beds to dry the influent air.

To remove harmful particulate material which could penetrate a granular charcoal bed, a filter is placed before the adsorbent. The filter medium may be made of paper, wool, cotton, asbestos, or other fibrous material. It must have the properties of high filtration efficiency for all particle sizes and low breathing resistance.

The adsorbent bed and filter combination are assembled in a metal container equipped with a valve which prevents exhaled air from entering the canister. Exhaled air would cause deterioration of the charcoal by reason of its high water vapor content.

Carriers for military masks are rugged canvas bags which can be carried conveniently by personnel and opened quickly for access to the mask.

Rebreathing masks. These contain chemical compounds such as sodium peroxide which react with carbon dioxide and water and liberate oxygen. These masks are worn in atmospheres that either are too heavily contaminated to be purified by a canister or are deficient in oxygen.

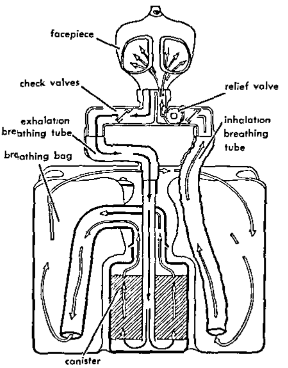


Fig. 2. The rebreathing mask.

Air-supply masks. These are used in similar situations. The air for these masks may come from a tank of compressed gas worn by the individual or may be delivered by a hose connected to a pump. All air-supply masks must be properly valved and regulated to prevent air pressure within the mask from becoming dangerous or uncomfortable.

Special-purpose masks. These include dust filters which cover only the nose and mouth and contain a filter but no charcoal, and oxygen masks for aviators which also cover only the nose and mouth and are essentially the same in operation as the air-supply mask. *See* ACTIVATED CARBON; CHEMICAL WARFARE. [S.D.S.]

Bibliography: P. Drinker and T. Hatch, *Industrial Dust*, 2d ed., 1954; A. A. Fries and C. J. West, *Chemical Warfare*, 1921; U.S. Department of the Army, *Protective Masks and Accessories*, Technical Manual TM 3–205, April, 1955.

Gas mechanics

The action of forces on gases. One of the simplest forces on a gas is illustrated by the gravitational attraction of the Earth for its atmosphere (*see* ESCAPE VELOCITY). From this and other observations, the weight of an ideal gas is found to be proportional to its molecular weight (*see* AVOGADRO'S LAW). Another mechanical property of a gas is its ability to transmit pressure in any direction (*see* PASCAL'S LAW). In this respect, a gas and a liquid behave similarly. *See* FLUID MECHANICS. [F.H.R.]

Gas tube

An electron tube that contains gas or vapor at low pressure in which an electrical discharge takes place. Gas tubes are of two general types: (1) cold-cathode tubes, in which the phenomenon known as glow discharge serves to maintain a con-

acterized by a relatively high voltage drop and a low current, while the hot-cathode arc tubes are characterized by a low voltage drop and relatively high current.

Volt-ampere characteristics. The two types of discharges may be most easily illustrated by means of the typical volt-ampere characteristics shown in Fig. 1. The portion OA represents the prebreakdown characteristic and is a function of electrode spacing. At A, which is the breakdown voltage, the nearly nonconducting gas ionizes and becomes a good conductor, and a self-sustained discharge is established. The portion BC is known as a normal glow. This glow is characterized by a nearly constant voltage, of the order of 80–400 volts (depending on gas and cathode material), for a range of current of the order of 1–100 milliamperes (ma). The portion CD, having a rising volt-ampere characteristic, is an abnormal glow. The characteristic EF having a negative volt-ampere characteristic is an arc discharge. For a given battery voltage V_b and circuit resistance R, two stable points exist. At P the discharge is a glow with current I_p while at Q the discharge is a low-voltage arc having a current I_Q. Both P and Q represent stable points of operation, and whether a glow-arc transition can occur depends on critical physical conditions at the cathode. The point E represents an unstable condition where operation is not possible for the indicated values of R and V_b.

The electron emission at the cathode in the glow region is largely due to positive-ion bombardment. The normal glow is characterized by a constant current density at the cathode surface. The normal glow occurs as long as current requirements of the external circuit are insufficient to cover completely the cathode surface with a glow, called the negative glow. At the point C the cathode becomes completely covered with the negative glow, and a further increase in current results in an increase in voltage, which is necessary to give the bombarding positive ions the energy required to produce the greater electron current required, giving the abnormal glow characteristic. Most of the voltage drop across the tube occurs between the cathode and the negative glow. The balance of the distance between the cathode and anode is filled with a glowing gas called a plasma, that requires only a small voltage to maintain its conducting state. *See* ELECTRICAL CONDUCTION IN GASES.

The voltage drop across an arc discharge is quite low compared with that of a glow discharge, being of the order of 10–20 volts. The low drop of the arc is due to an efficient mechanism of electron emission from the cathode and to the presence of large numbers of positive ions that neutralize the space charge of the electrons.

Cold-cathode gas tubes. Cold-cathode gas tubes, or glow tubes, require no cathode heater power and are therefore always ready for instant service. Glow tubes are commonly constructed as 2-element (diode) or 3-element (triode) tubes.

Cold-cathode diode. This type of tube is usually either a glow lamp or a voltage-regulator tube.

The glow lamp has two electrodes, usually of equal size. The light from the negative glow makes the tube useful as an indicator lamp. These lamps

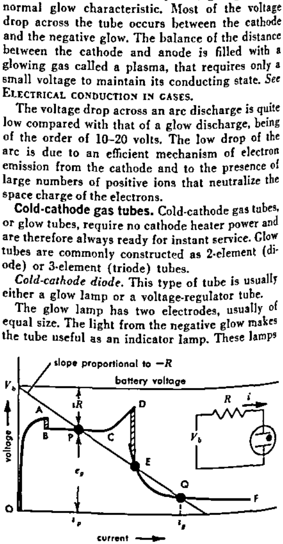


Fig. 1. Typical volt-ampere characteristics of a cold-cathode gas tube.

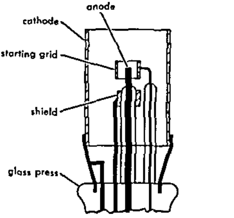


Fig. 2. The cold-cathode gas triode.

have a small series resistance built in the base and are designed to operate directly on 120 or 240 volts. Similar tubes are also available with special bases that have no built-in resistor and are used in circuits containing their own ballast resistor. *See* NEON GLOW LAMP.

The voltage-regulator tube operates in the normal glow region and therefore has a nearly constant anode-to-cathode voltage drop over a range of current. These tubes are used to maintain a dc potential substantially constant for adjustments of input voltage or load current. *See* VOLTAGE-REGULATOR TUBE.

Cold-cathode triode. This tube is sometimes called a grid-glow tube. Tubes of this type have a third starting electrode, which surrounds the anode

positive voltage to break down the cathode-starter gap, after which the discharge transfers to the main anode. The anode potential need be no higher than that necessary to sustain the discharge. A typical tube of this type has a maximum anode voltage drop of 85 volts, average current of 25 ma, and a starter breakdown voltage in the range 73–105 volts.

Hot-cathode gas tubes. These low-pressure arc-discharge devices operate with the characteristic EF of Fig. 1. Three representative types may be distinguished, (1) the Tungar (sometimes called a Rectigon); (2) the phanotron; and (3) the thyratron. The first two are simple rectifier tubes while the third is a control tube having one or two grids between cathode and anode. Both phanotrons and thyratrons may be built with glass, ceramic, or metal envelopes.

Tungar tube. This tube is a low-voltage rectifier having a hot thoriated-tungsten filament, an anode

ment and the argon gas. This tube has a low arc drop of the order of 8 volts, and a current range of 1–15 amperes, which is well suited to battery charging service. The relatively high gas pressure permits higher than normal operating temperature of the filament by suppressing filament evaporation. Such atoms of tungsten and thorium as are evaporated are largely ionized near the cathode and returned to the filament so that a long life is obtained. However, the high pressure results in slow deionization of the gas. During inverse voltage periods (the portion of the cycle in which the anode is negative), a glow may be started at voltages of 200–300 volts with consequent danger of the glow-arc transition (*see* ARC DISCHARGE). The tube starts at a relatively low forward voltage due to the copious supply of electrons available and to the low gas density in the vicinity of the cathode where the gas temperature is relatively high.

Phanotron tube. The phanotron is a thermionic (hot-filament) diode rectifier tube utilizing an arc discharge in mercury vapor or an inert gas, usually xenon. Mercury is the preferred filling where a low arc drop is desired, whereas one of the inert gases is used where independence of ambient temperatures over a wide range is desired. The arc drop for the mercury tube is about 10 volts and that for a xenon tube is about 12 volts. In both cases the arc drop is about equal to the ionization potential of the filling gas or vapor. The inert-gas tubes suffer from one defect not present in mercury tubes, namely, the gas pressure eventually is reduced to a low value and the tube becomes inoperative. The gas pressures in the hot-cathode arc tubes are of the order of 6×10^{-3} mm Hg, which for a mercury tube corresponds to a temperature of about 40°C at the coldest point. The point at which mercury condenses determines the average vapor pressure in the tube. At gas pressures of this order the Paschen breakdown curve for the electrode spacing usually found in tubes has a steeply increasing voltage relation as the pressure is decreased. The arc tubes are designed so that the product of gas pressure times separation between electrodes is such that, in the absence of electron emission, a high breakdown voltage exists. This is the condition existing during the inverse voltage period of a rectifier when the anode is negative. Thus these low-pressure arc tubes may be used to rectify relatively high voltages. When electron emission is present, as at the cathode during a normal period for conduction (when the cathode is negative and the anode positive), the arc discharge is easily started. When the forward voltage increases to a value of the order of the ionization potential of the gas, or vapor filling, some electrons gain sufficient energy to ionize when they collide with the gas atoms. This results in one or more electron avalanches and breakdown of the gas is effected.

In a high-vacuum rectifier only those electrons directly acted on by electrostatic field lines from the anode can be utilized. The presence of gas in the arc tubes permits the utilization of e' emitted in rather deep cavities in the cathode ture, so a high value of electron emissi

can be obtained. This is true because electrons emitted thermionically deep in a cavity will undergo many collisions with gas atoms with resulting changes in direction so that they can reach the main discharge path. Figure 3 shows the structure of such a cavity cathode. The surfaces of the vanes, the outer side of inner cylinder, and the inner side of outer cylinder are coated with an oxide type of emitter (*see* THERMIONIC EMISSION). The emitting coating is usually a mixture of barium and strontium carbonates reduced to the oxides. The maximum current that can be drawn from the cathode is the temperature-limited, or saturation, current. If the circuit conditions require a higher current than this, the arc drop will increase, largely at the cathode, and the energy of the bombarding positive ions will be increased so as to cause the required increase in electron emission. This will result in destruction of the oxide cathode. The critical value of voltage producing this destructive effect is about 25 volts for mercury-vapor and inert-gas tubes. Such overloads should be avoided.

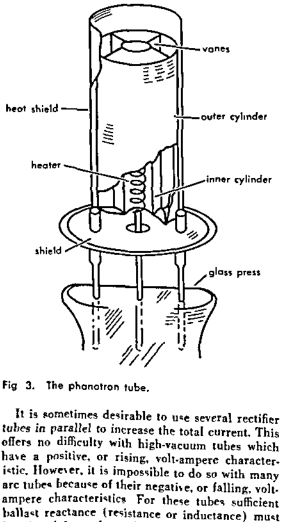


Fig 3. The phanotron tube.

It is sometimes desirable to use several rectifier *tubes in parallel* to increase the total current. This offers no difficulty with high-vacuum tubes which have a positive, or rising, volt-ampere characteristic. However, it is impossible to do so with many arc tubes because of their negative, or falling, volt-ampere characteristics For these tubes sufficient ballast reactance (resistance or inductance) must be placed in each anode lead to make the combined arc-drop and reactance-voltage characteristic positive. Typical phanotrons have average current ratings from 1.25–10 amperes and peak inverse voltage ratings of 10,000–22,000 volts.

Thyratron tube. The thyratron is a hot-cathode gas tube containing one or more grid elements between the cathode and anode. It is an important current-control tube, having many applications. *See* THYRATRON.

Noise, oscillations and surges. Hot-cathode gas tubes generate both electrical noise and oscillations that are characteristic of the tube construction and of the gas. The noise spectrum is substantially flat from frequencies of 25 cps to 1 megacycle per second or higher. One or more fundamental modes of oscillation at fairly discrete frequencies are usually present, often with many harmonics for each fundamental mode. The peak-to-peak variational component of the anode, or arc, voltage is of the order of the ionization potential of the gas or vapor present. The rms value of the noise voltage is of the order of several volts. The presence of these noise voltages in the anode potential of a gas tube during its conducting period is sometimes the source of considerable disturbance in nearby amplifiers, receivers, and other sensitive circuits, either by direct electrical coupling or by electromagnetic radiation. Usually the unwanted signal can be bypassed or shielded when its source is recognized. Hot-cathode thyratrons are sometimes used as noise-source generators when placed in suitably selected and oriented magnetic fields. The coherent oscillations of many tubes can be suppressed by the use of suitably oriented magnetic fields. *See* NOISE, ELECTRICAL.

When the arc current in a thyratron is increased above a critical value, any constriction of the arc path, such as the holes in the grids, becomes overloaded and serious voltage surges result. This phenomenon is caused by the complete ionization of the gas in the grid holes. When this occurs, the positive ions produced are driven out of the grid area, leaving a "vacuum" zone, that is, gas starvation results. The sudden stopping of current produces very high voltages in associated inductances. [J.D.C.]

Bibliography: J. D. Cobine, *Gaseous Conductors*, 1958; J. D. Cobine and C. J. Gallagher, Noise and oscillations in hot-cathode arcs, *J. Franklin Inst.* 243:41–54, 1947; J. D. Cobine and C. J. Gallagher, Noise in gas tubes, *Electronics*, 20(3):144, 1947; J. D. Cobine and C. J. Gallagher, Effects of magnetic field on oscillations and noise in hot-cathode arcs, *J. Appl. Phys.*, 18(1):110–116, 1947; W. D. Cockrell (ed.), *Industrial Electronics Handbook*; 1958; G. N. Glasoe and J. V. Lebacqz (eds.), *Pulse Generators*, vol. 5, 1948; A. W. Hull and F. R. Elder, The cause of high-voltage surges in rectifier circuits, *J. Appl. Phys.*, 13(6):372–377, 1942; L. Malter and E. O. Johnson, Studies of thyratron behavior, *RCA Rev.*, 11(2):165–189, 1950; L. Malter, E. O. Johnson, and W. M. Webster, Studies of externally heated hot-cathode arcs, *RCA Rev.*, 12(3):415–435, pt. 1, 1951; L. Tonks, Theory and phenomena of high current densities

in low-pressure arcs, *Trans. Electrochem. Soc.*, 72:167–182, 1937.

Gas turbine

A heat engine that converts heat into work using gas as the working medium, and that delivers its mechanical output as a shaft rotation.

Gas turbine cycle. In the usual gas turbine, the sequence of thermodynamic processes, in which the working fluid is eventually returned to its original state, consists basically of compression, addition of heat in a combustor, and expansion through a turbine. The flow of gas during these thermodynamic changes is continuous in the basic, simple, open-cycle arrangement (Fig. 1*a*).

This basic open cycle can be modified through the addition of heat exchangers and multiple components for reasons of efficiency, power output, and operating characteristics (Fig. 1*b*–*d*). A regenerator recovers exhaust heat and returns it to the cycle by heating the air after compression and before it enters the combustor. An intercooler reduces the work of compression by removing the heat of compression. A second stage of heating can be added between sections of the turbine. In various combinations, the auxiliary features provide means for meeting a wide range of operating needs.

Gas turbine types. The various gas turbine cycle arrangements can be operated as open, closed, or semiclosed types.

Open cycle. In the open-cycle gas turbine, there is no recirculation of working medium within the structural confines of the power plant, the inlet and exhaust being open to the atmosphere (Fig. 1). This cycle offers the advantage of a simple control and sealing system. It also can be designed for high power-to-weight ratios (aircraft units) and for operation without cooling water. Most gas turbine plants are of this type.

Closed cycle. In the closed-cycle gas turbine, essentially all the working medium (except for seal leakage, bleed loss and any addition or extraction of working medium for control purposes) is continuously recycled (Fig. 2*a*). Heat from a source such as fossil fuel or nuclear reaction is transferred through the walls of a closed heater to the cycle. The closed cycle can be charged with gases other than air such as helium, carbon dioxide or nitrogen. This is a particular advantage with a nuclear heat source. The other advantages of the closed cycle

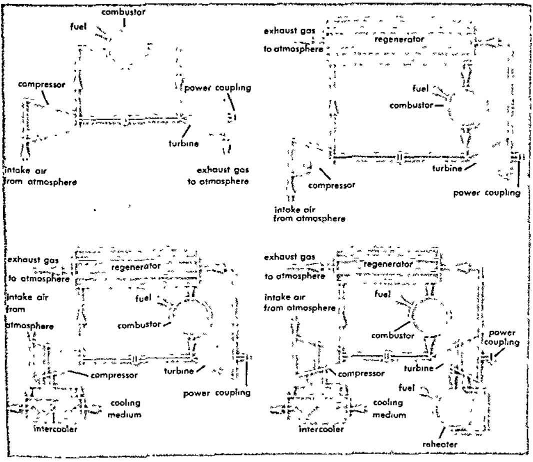


Fig. 1. Series-flow, single-shaft, gas turbine power plant. (*a*) Simple open cycle. (*b*) Open cycle with regenerator (*c*) Open cycle with intercooler and regenerator. (*d*) Open cycle with intercooler, regenerator, and reheater.

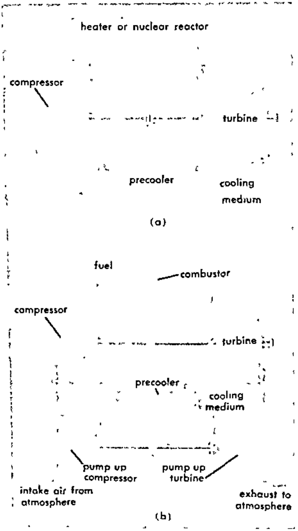


Fig. 2. Gas turbine power plant. (a) Closed cycle with precooler, series flow, single shaft. (b) Semiclosed cycle with precooler, series flow, single shaft.

are (1) clean working fluid, (2) control of the pressure and composition of the working fluid, (3) high, absolute pressure and density of the working fluid, and (4) constant efficiency over wide load range. A precooler is required to reduce the temperature of the working fluid before recompression. The higher densities of the working fluid increase the horsepower capacity of a plant of given volume. Changing the absolute pressure level at the compressor inlet changes the weight of working fluid circulated without changing the compression ratio or the temperatures, which results in relatively constant efficiency over a wide load range. The major disadvantage of the closed-cycle gas turbine plant is the cost and size of the required high-temperature heater.

Semiclosed cycle. In the semiclosed cycle gas turbine, a portion of the working fluid is recirculated (Fig. 2*b*). This type requires a precooler for the recirculated gas, and a charging compressor to provide the necessary air for combustion. The semiclosed cycle can operate at high densities. The major disadvantages of this cycle are the corrosion and fouling, which occur with the recirculation of the products of combustion, particularly when the fuels used have high sulfur or ash content.

Generalized performance. The over-all performance of a given gas turbine power plant cycle depends basically on component efficiencies, pressure and leakage losses, pressure ratio (ratio of the highest to the lowest pressure), and temperature level.

For the simple cycle, the influence of pressure ratio on performance is illustrated in Fig. 3*a*. The curves are based on ambient inlet conditions of 80°F and 1000 ft altitude, 85% compressor efficiency, 90% turbine efficiency, 95% combustion efficiency and 5% combustor pressure loss.

The effect of pressure ratio on plant efficiency for the four basic gas turbine cycle arrangements

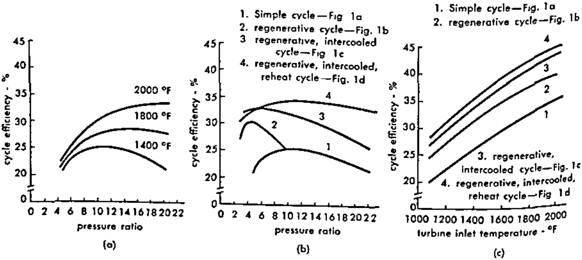


Fig. 3. Effect of pressure ratio on thermal efficiency. (a) Simple gas turbine cycle at various turbine-inlet temperatures. (b) Various gas turbine cycles at constant turbine-inlet temperature. (c) Effect of turbine-inlet temperature on thermal efficiency for various gas turbine cycles at optimum pressure ratios

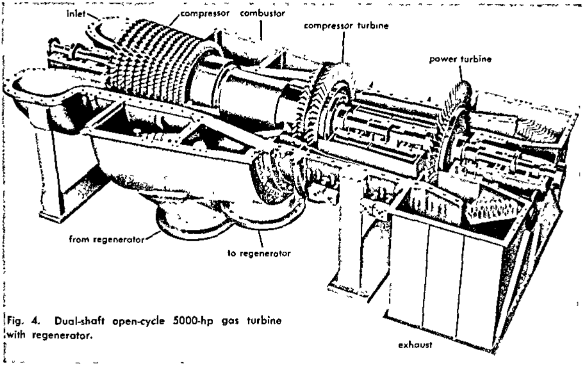


Fig. 4. Dual-shaft open-cycle 5000-hp gas turbine with regenerator.

illustrated in Fig. 1 is shown in Fig. 3*b*; the effect of inlet temperature on cycle efficiency for the four basic cycles is shown in Fig. 3*c*. The curves are drawn for the optimum pressure ratio for each cycle arrangement at the various turbine inlet temperatures.

Gas turbine components. To achieve such overall performance, each process is carried out in the engine by a specialized component (Fig. 4). Air for the combustion chamber is forced into the engine by a compressor. In an aircraft, the intake may advance into the air fast enough to ram air into the engine. Fuel is mixed with the compressed air and burned in combustors. The heat energy thus released is converted by the turbine proper into rotary energy. Because of the high initial temperature of the combustion products, excess air is used to cool the combustion products to the allowable turbine inlet design temperature. To improve efficiency, heat exchangers can be added on the gas turbine exhaust to recover heat energy and to return it to the working medium after compression and prior to its combustion.

Compressors. Two basic types of compressors are used in gas turbines: axial, and centrifugal. In a few special cases a combination type known as a mixed wheel, which is partially centrifugal and partially axial, has been used. The axial-flow compressor is the most widely used because of its ability to handle large volumes of air at high efficiency. For small gas turbines in the range of 500 hp and less, the centrifugal replaces the axial since it has comparable efficiency when handling reduced volume flow, and is smaller and more compact.

Axial and centrifugal compressors both have a stall or pumping limit where flow reverses. This limit is usually encountered on starting with stationary power units, and at high altitude and high speed with aircraft units. This stall limit is shifted out of the operating range of the gas turbine by the use of compressor bleed, variable compressor vanes, or water injection. All three methods are applied so as to reduce the aerodynamic loading on the stalled compressor stages.

Ram effect Aircraft gas turbines moving at high speeds obtain a pressure rise from the ram effect in addition to the compressor pressure rise. The ram effect is the recovery of part of the air velocity, due to the forward motion of the plane, and conversion of this velocity energy to pressure.

Combustors. Combustors, sometimes referred to as combustion chambers, for gas turbines take a wide variety of shapes and forms. All contain fuel nozzles to introduce and meter the fuel to the gas stream and to atomize or break up the fuel stream for efficient combustion (Fig. 5). All are designed for a recirculating flow condition in the region of the nozzle so as to develop a self-sustaining flame front in which the gas speed is lower than the flame propagation speed. In addition to being designed to

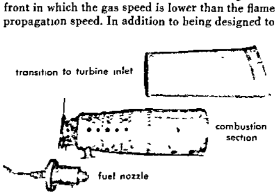


Fig. 5. Gas turbine combustor components.

burn the fuel efficiently they also uniformly mix excess air with the products of combustion to maintain a uniform turbine-inlet temperature.

The combustor must bring the gas to a controlled, uniform temperature with a minimum of impurities and a minimum loss of pressure. In the open-type gas turbine, the large excess of air must be controlled to avoid chilling the flame before complete combustion has taken place. Extremely high rates of combustion are common, the heat release being 1,000,000–5,000,000 Btu per cu ft per hour per atmosphere or from 5 to 20 times that of high-output, steam-boiler furnaces. Refractory linings are not suitable since they cannot withstand the vibrations and velocities. Metal liners and baffles cooled by the incoming air are commonly used.

Turbine wheels. Two types of gas turbine wheels are used: radial-inflow, and axial-flow. Small gas turbines use the radial-flow wheel. For large volume flows, axial turbine wheels are used almost exclusively. Although some of the turbines used in the small gas turbine plants are of the simple impulse type, most high performance turbines are neither pure impulse nor pure reaction (*see* TURBINE). The high performance turbines are normally designed for varying amounts of reaction and impulse to give optimum performance. This usually results in a blade which is largely impulse at the hub diameter and almost pure reaction at the tip.

Turbine cooling. Gas turbines all employ cooling to various extents and use a liquid or gas coolant to reduce the temperature of the metal parts. The cooling systems vary from the simplest form where only first-stage disk cooling is involved to the more complex systems where the complete turbine (rotor, stator, and blading) is cooled. At high temperatures of operation the turbine material is designed to have a predetermined life, which can vary from a few minutes for missile and aircraft applications to 100,000 hours or more for industrial units.

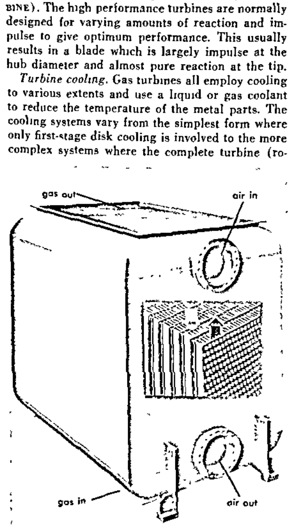


Fig 6. Air-to-gas regenerator, partly cut away.

Heat exchangers. Two basic types of heat exchangers are used in gas turbines: gas-to-gas, and gas-to-liquid. An example of the gas-to-gas type is the regenerator, which transfers heat from the turbine exhaust to the air leaving the compressor. The regenerator must withstand rapid large temperature changes and must have low pressure drop. Regenerators of both the shell-and-tube and the extended-surface type are used (Fig. 6). Rotary regenerators having high performance and reduced weight are under development. Intercoolers, which are used between stages of compression, are air-to-liquid units. They reduce the work of compression, and the final compressor discharge temperature When used with a regenerator, they increase both the capacity and efficiency of a gas turbine power plant of a given size.

Gas turbine arrangements. Gas turbines can be constructed for single- or multiple-shaft arrangements. They can be arranged to supply power, high-pressure air, or hot-exhaust gases either singly or in combination.

A single-shaft unit, consisting of a compressor, combustor and turbine, is a compact, lightweight power plant. It is capable of rapid starting and loading and has no standby losses. It can be arranged to use little or no cooling water, which makes it particularly attractive for powering transportation equipment and as a mobile standby and emergency power plant. This type of plant can compete efficiently with small steam plants. Its simplicity involves a minimum of station operating personnel; some of these plants are arranged for completely remote operation.

To improve efficiency, the energy in the exhaust gases can be used either in a waste-heat boiler, or in combination with other processes. For example, a unit arranged to supply process air at 35 psig for operation of blast furnaces can use blast-furnace gas for its fuel (Fig. 7*a*).

Gas turbines can also be arranged to use a nuclear heat source. For aircraft application, the cycle is open because of weight and space considerations. The combustor is replaced with a nuclear reactor (Fig. 2*a*). For shipboard and stationary units, the cycle is closed so that fission products can be contained within the power plant in the event of a nuclear accident or fuel element failure. A closed-cycle, gas turbine power plant for marine service can operate directly with a high-temperature, gas-cooled reactor (Fig. 7*b*).

Gas turbine fuels. In the open-cycle plant, prod ... h the ... This ... stion

are relatively free of corrosive ash and of residual solids that could erode or deposit on the engine surfaces. Natural gas, refinery gas, blast-furnace

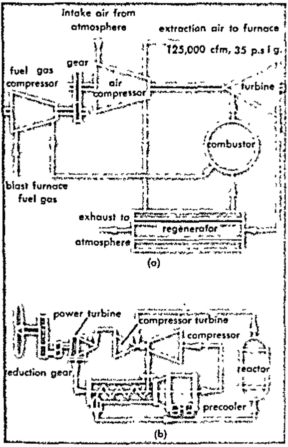


Fig. 7. Cycle arrangements. (a) For 125,000 cfm extraction unit. (b) For closed-cycle gas turbine with gas-cooled reactor.

gas, and distillate oil have proved to be ideal fuels for open-cycle gas turbines. Combustion chamber and fuel nozzle maintenance are negligible with these types of fuel. Residual fuel oil treated to avoid hot ash corrosion and deposition is also satisfactory although it requires frequent cleaning of fuel nozzles and higher combustor maintenance. Vanadium pentoxide and sodium sulfate are the principle ashes that have been found to cause corrosion and deposition in the 1250°F to 1600°F temperature ranges of modern-day gas turbines. The treatment of residual oil to make it suitable consists of washing with water to remove sodium and introducing additives to raise the fusion temperature of the ash. Development continues on techniques for using solid fuel such as coal and peat. The major problems in using coal and peat for open-cycle gas turbines lie in the separation and elimination of the fly ash before it passes through the turbine where it would quickly erode the turbine blading.

In closed-cycle turbines, gaseous fuel, distillate oils, and coal are satisfactory. Residual fuels also must be treated to avoid corrosion in the heater parts. Ash erosion is not a major problem because of the lower velocity of the combustion gases over the heater surfaces.

Fuels for semiclosed cycle turbines must be selected to minimize both erosion and corrosion. Since the products of combustion are circulated through the entire cycle, the fuel must be selected to avoid corrosion at all pressures and temperatures of the cycle. These requirements limit the use of the semiclosed plant to fuels of low ash and sulfur content.

Gas turbine applications. Gas turbine power plants have been successfully applied in the following industries where their characteristics have proved superior to competitive power plants.

Aviation. In the propulsion of aircraft, the gas turbine power plant is rapidly supplanting the reciprocating engine in large, high-speed airplanes as a turboprop or turbojet engine. This change is due primarily to its very high power-to-weight ratio and its ability to be built in large horsepower sizes with high ratio of thrust per frontal area (Fig. 8). It makes use of the ram effect on the compressor inlet to give almost constant thrust at all aircraft speeds. *See* AIRCRAFT ENGINE PERFORMANCE.

Gas pipeline transmission. Gas turbines have been installed along pipelines to drive centrifugal compressors. Turbine sizes range from 1800 hp to 14,000 hp. Gas fuel is normally used, though units have been arranged for dual fuel firing using either natural gas or distillate oil.

Petroleum. In the petroleum industry the gas turbine generates power, drives air and gas compressors in refineries, supplies extraction air or exhaust gases for process, and in the oil fields drives gas compressors to maintain well pressure.

Steel. In the steel industry gas turbines are used to drive air compressors, to generate electric power, and to furnish extraction air for processes such as blast-furnace operation, as mentioned above (Fig. 9*a*). The primary fuel in these applications is blast-furnace gas with a number of units arranged for dual fuel firing.

Marine. Prototype gas turbines are being tested for marine use in ratings up to 6000 ship horsepower (shp). These units operate with residual oil

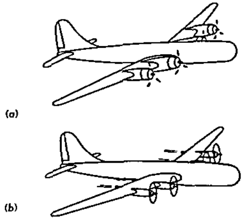


Fig. 8. A reciprocating engine (*a*) has greater frontal area for a given horsepower than an [illegible] tion gas turbine engine (*b*).

as fuel processed through fuel treatment equipment. A regenerative-cycle gas turbine with a separate power turbine provides the wide speed range required by such applications as marine service (Fig. 4).

Marine installations of gas turbines range from small units used for propulsion of short-range, high-speed craft, emergency generator drive, minesweep boat propulsion and generator drive, deicing, smoke generation, fire-pump drive, and pneumatic power applications, to large boost engines used in combination with steam turbines.

Electric utilities. Gas turbines perform a variety of functions in electric power generation. Peaking service uses simple single-shaft units in sizes up to 25,000 kw. Gas turbines have been installed as hydro-standby. The largest station of this type is the 40,000 kw Beznau, Switzerland power plant which has two units in operation, one rated 27,000 kw, the other 13,000 kw. Rail-mounted mobile power plants provide 5,000 to 6,000 kw for emergency service. The compactness and simplicity of the gas turbine make it possible to house these units in a single cab. Units from 5,000 to 30,000 kw carry base load. These units have been simple cycle machines in low fuel cost areas; units with regenerative, intercooled cycles are used in high fuel cost areas.

Combined steam and gas plants. Efficiency of a power plant is improved by combining a gas turbine with a steam turbine. Various combinations are possible; gas turbine exhaust can heat feed water for the steam turbine, fuel-fired boilers can use gas turbine exhaust as combustion air (Fig. 9*a*), or, in supercharged boiler plants, the high-pressure boiler can serve as the pressurized combustor for the gas turbine (Fig. 9*b*). The gain in efficiency by using turbine exhaust (Fig. 9*a*) depends on efficiency of the over-all steam plant to which the combined cycle is compared. For the most efficient plant using the best steam conditions and largest number of feedwater heaters, a gain of 2% is possible, while for the less efficient plant a gain of 5% is possible.

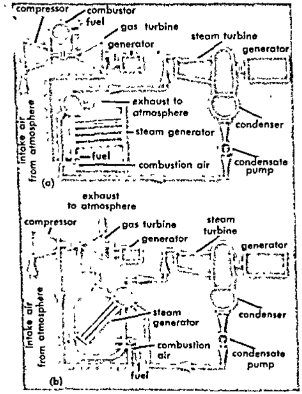


Fig. 9. Cycle arrangements. (a) For combined steam and gas turbine cycle with gas turbine, exhaust-fired steam generator. (b) For combined steam and gas turbine cycle utilizing a pressurized-steam generator.

Where the steam generator is pressurized with air from the gas turbine compressor and the hot gases from th[illegible] expanded through the gas turbi[illegible] ciency of 4–8 [illegible] used for comparison. These improvements can be made to all steam plant cycles including the most efficient currently in operation, which employ the super-pressure cycle. [T.J.P.]

Bibliography: D. G. Shepherd, *An Introduction to the Gas Turbine*, 1949; H. A. Sorensen, *Gas Turbines*, 1951.

Gasket

A gasket is used to make a pressure-tight joint between stationary parts, such as cylinder head and cylinder, that may require occasional separation. Gaskets are known as static seals, as compared with packing or dynamic seals (*see* SEAL, PRESSURE). In packings the parts are frequently in motion, as in piston rods and valve stems.

Gaskets are made of sheet materials such as natural or synthetic rubber, cork, vegetable fiber such as paper, asbestos and plastic pastes, or of soft metallic materials such as lead and copper. Rubber in the form of O-rings is used for light pressure. [P.H.B.]

Bibliography: W. Kent, *Mechanical Engineer's Handbook*, vol. 1, 12th ed., 1950; Society of Automotive Engineers, *SAE Handbook*, revised annually.

Gasoline

A motor fuel consisting largely of hydrocarbons. It is derived from crude petroleum by a number of different processes. The names of the processes and the products formed are as follows: condensation, casing-head gasoline; absorption, natural gasoline; distillation, straight-run gasoline; alkylation, alkylate; polymerization, polymer gasoline; reforming, reformate; thermal cracking, thermal gasoline; and catalytic cracking, cat-cracked gasoline. *See* PETROLEUM PROCESSING.

Gasoline of commerce consists of a mixture of these materials in various proportions, plus small quantities of additives. Gasoline may also be made from shale oil, tar sands, gilsonite, and coal and by the Fischer-Tropsch process ($CO + H_2$) and may include alcohols and aromatics from coal. The total volume from these sources is insignificant

compared with that from crude oil. Gasoline is utilized by internal-combustion engines such as those in automobiles and airplanes. The gasoline is mixed with an approximately stoichiometric quantity (13 to 1 by weight) of air, usually by means of a carburetor; the mixture is then compressed in the combustion chamber about 5- to 10-fold and ignited by means of a spark. *See* INTERNAL COMBUSTION ENGINE.

Motor gasoline. Gasoline for automobile engines has an ASTM (American Society for Testing Materials) boiling range of about 80–430°F (containing hydrocarbons in the range C_4–C_{12}) and in the United States is usually sold in two grades differing in octane number and price. The premium grade has an octane number 7–10 units greater than that of the regular grade. During the 1950s some companies introduced a third grade, super premium, having a research octane number above 100. The physical and chemical properties of gasoline are controlled by specifications designating the boiling range, volatility, octane number, stability, and amount of minor constituents. Specifications have been published by the ASTM, the IP (Institute of Petroleum), and the Armed Forces, but each oil company also has its own specifications for each marketing area. The octane number of the average (pool) gasoline in the United States has increased steadily over the years as the compression ratio and horsepower of engines have increased. The rest of the world shows a similar trend but at lower levels. An increase in the compression ratio causes an increase in the octane requirement of the engine because of the fact that the temperature and pressure of the charge are increased on compression. After extensive engine operation, the octane requirement becomes greater because of the accumulation of deposits in the combustion chamber. The deposits increase the compression ratio because of their bulk, the temperature because of their insulating properties, and the oxidative reaction rate because of their catalytic effect on combustion. The octane-requirement increase varies with individual engines but is in the range of about 5–10 octane numbers at equilibrium (after driving a new or cleaned-up engine for 2000–3000 miles). The octane number is usually expressed in terms of the research or F-1 value determined in the CFR (Cooperative Fuel Research) engine at 600 rpm with a mixture temperature of 125°F (ASTM D-357). It may also be determined by the motor or F-2 method at the more severe condition of 900 rpm and 300°F mixture temperature (ASTM D-908). The F-2 value is generally lower than the F-1 by about 6–14 numbers. This difference is known as the sensitivity. The actual octane requirement of an automotive engine depends on the operating conditions and is usually highest under high load conditions. It decreases by about one to two octane numbers per 1000 ft with increasing elevation because of decreasing charge density.

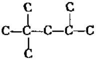
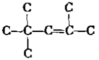
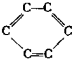
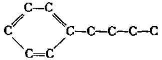
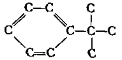
The octane number of a gasoline depends upon the type of crude used, the processing techniques employed, and the amount of antiknock agents added, mainly tetraethyllead (TEL). The octane number of gasoline-blending components generally increases in the order straight-run (SR), thermally cracked (TC), cat-cracked (CC). This is because of the structure of the hydrocarbons making up these gasolines. The octane number of SR or cracked stocks can be increased by catalytic reforming (CR), generally over a platinum-containing catalyst. Branching and increasing unsaturation generally increases octane number as shown in Table 1. Because of their mode of manufacture the SR, TC, CC, and CR components vary in their hydrocarbon composition. Table 2 gives the compositions of the components prepared from West Texas crude (fractions from other sources are apt to be quite different in composition, particularly the SR and TC).

Sulfur compounds are generally present in minor amounts. They are undesirable because the sulfur is oxidized to sulfur dioxide and trioxide in the combustion chamber. This may result in the formation of corrosive sulfurous and sulfuric acids in the crankcase. For this reason, and because of the generally deleterious effect of sulfur compounds on stability and TEL susceptibility, the sulfur content is rarely allowed to exceed 0.25% wt. The sulfur compounds consist largely of disulfides, sulfides, and thiophenes. Mercaptans are generally extracted with caustic or solutizer solutions, or converted to disulfides (doctor or copper treating), or to sulfides (inhibitor sweetening). The nitrogen compounds are largely pyrroles, pyridines, and (in CC only) anilines. The nitrogen content is generally less than 0.02% wt. The chief types of oxygen compounds are alkyl phenols (0–0.1% wt) and carboxylic acids (0–0.1% wt). They are generally largely removed during caustic treating.

Volatility is important in order to ensure sufficient hydrocarbon vapor for ignition. It is adjusted to be greater in winter than in summer by adding extra butanes or pentanes. This also affects the flash point, which is the temperature at which the gasoline will ignite in the presence of an open flame. Values are about −50° for winter and −40°F for summer gasolines.

Aviation gasoline. Aviation gasoline is generally composed of alkylate mixed with specially selected or converted fractions of straight-run gasoline. Relatively pure hydrocarbons such as toluene and isopentane are often included in the blend. Various grades are marketed (values of 100 and below are octane numbers, above 100 are performance numbers). The grades are characterized by two numbers representing ratings for lean cruise (ASTM D-614, aviation, F-3) and rich mixture (ASTM D-909, supercharge, F-4). The latter may be as much as 30 numbers higher than the former, for example, 115–145. The higher grades are generally utilized by large commercial and military engines, the lower grades by private planes. Although the compression ratio of the large aircraft engines is not high (only about 5–7), their octane requi-

Table 1. Relation between molecular structure and octane number

Compound	Carbon skeleton	Octane number F-1	Octane number F-2
n-Heptane	C—C—C—C—C—C—C	0	0
n-Hexane	C—C—C—C—C—C	21.8	26 0
2,2,4-Trimethylpentane (isooctane)		100	100
1-Hexene	C=C—C—C—C—C	76.4	63.4
Diisobutylene		105.3	88.6
Benzene		>120	115
n-Butylbenzene		102	95.3
tert-Butylbenzene		>115	108

ment is high because the pressure of the initial charge is increased by supercharging. The compression ratio of engines on most small private planes is about 8–9, but they are not usually supercharged.

Because of the nonolefinic nature of the hydrocarbons in aviation gasoline, it is very resistant to oxidation during storage, and antioxidants are incorporated chiefly to prevent oxidation of the tetraethyllead. For this, the trialkylphenol type is preferred (a maximum of 4.2 lb/1000 bbl is permitted).

Oxidation stability. During storage, gasoline is subject to oxidation as a result of the attack of oxygen on reactive hydrocarbon, nitrogen, and sulfur compounds, which results in the formation of gum and peroxides. The N and S compounds are more reactive than the average hydrocarbon as indicated by the fact that the gum formed is always richer in nitrogen and sulfur than is the substrate. The oxidation is a chain reaction and proceeds through the formation of free radicals of the type R·, ROO·, and HO·. Usually it exhibits an induction period, that is, the reaction rate is quite low for a period of time and then suddenly accelerates. At ambient temperature, the induction period may last 6–12 months, whereas at 100°C it is about 6–12 hours. The most important product of oxidation is the gasoline gum which is an oxygenated dimer or trimer with an average composition C = 72%, H = 10%, S = 0.5%, N = 0.2%, O = 17% and a molecular weight of about 400. High concentrations may cause difficulty in engine operation because of the formation of deposits on the carburetor, induction manifold, valves, and combustion-chamber. Peroxides which are formed during oxidation decrease the octane number of the gasoline.

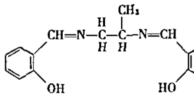
Gasoline additives. The formation of gum can be reduced and the induction period extended by the addition of chemical compounds known as gum inhibitors which consist of antioxidants and metal deactivators. Commercial antioxidants are substituted aromatic amines and phenols, the most important being *N,N'*-di-*sec*-butyl-*p*-phenylenediamine, 2,6-di-*tert*-butyl-4-methylphenol, 2,4-dimethyl-6-*tert*-butylphenol, *N*-*n*-butyl-*p*-aminophenol. Antioxidants function by reacting with the free radicals formed as intermediates in the oxidation reaction, thus stopping the chain reaction. Good antioxidants must be reactive enough to undergo this type of reaction but not so reactive that they are readily oxidized and hence removed early in the course of the reaction. The purpose of the substituent alkyl groups is to shield or hinder the reactive —OH and —NH groups to the proper degree to provide optimum protection for the gasoline. Antioxidants are generally added to the extent of 10–30 ppm; this amount may extend the induction period by 5- to 10-fold at ambient conditions.

Metal deactivators are chelating compounds which reduce the rate of gum formation by nullifying the catalytic effect of active metals such as cop-

Table 2. Compositions of gasolines

Component	Straight run	Thermally cracked	Catalytically cracked	Catalytically reformed
Paraffins, % vol	58	25	26	29
Naphthenes, % vol	34	19	12	4
Olefins, % vol	0	43	27	0
Diolefins, % vol	0	3	<1	0
Aromatics, % vol	8	11	35	67
Octane number (F-1)	71	72	90	97
Sulfur, % wt	0.07	0.24	0.1	(0.01)
Nitrogen, % wt	(0.02)	(0.05)	(0.01)	(0.001)
ASTM boiling range, °F	130–229	162–357	90–438	174–413

per and iron by forming coordination compounds (chelates) with them. The catalytic effect of copper is such that as little as 1 ppm can increase the rate of gum formation by several-fold. Although many compounds are capable of chelating with metals, the only one of commercial significance in the United States is disalicylal-1,2-propane diimine. The corresponding ethylenediamine



derivative is also used to some extent abroad. Generally, the addition of 1.5 times the stoichiometric amount of copper present in the gasoline is recommended. The usual dosage is about 5 ppm. It is particularly important to add deactivator to a copper-sweetened gasoline.

Dyes are added to the extent of about 5 ppm for identification and advertising purposes.

The most important antiknock is tetraethyllead added to the extent of about 1–4 ml per gallon (0.03–0.13% wt Pb) in motor gasoline, 4.6 ml in aviation. *See* ANTIKNOCK AGENTS; TETRAETHYLLEAD.

Ethylene dichloride and ethylene dibromide are added, as part of the TEL concentrate, to the extent of 0.5–1.5 ml per gallon to reduce the accumulation of lead deposits in the combustion chambers.

Agents added to prevent carburetor icing are of two types, alcohols (isopropyl) 1–2% vol and surfactants, 50–100 ppm.

Surfactant type materials may be added to the extent of 10–50 ppm to reduce corrosion in the fuel system.

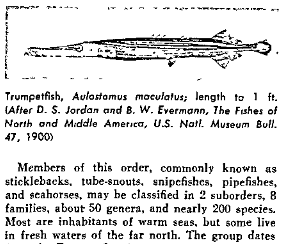
Deposit modifiers, generally phosphorus compounds, such as aryl phosphates, are added to ensure that the lead left in the combustion chamber is predominantly in the form of lead phosphate. They are added to the extent of 20–30% wt of the lead. They reduce preignition and spark plug fouling. Boron compounds are also used at low (generally < 0.01% wt B) concentrations to reduce preignition and octane-requirement increase but generally do not reduce spark-plug fouling.

Detergents may be added to the extent of a few hundred ppm to reduce carburetor malfunction.

A light lubricating oil (0.1–0.3%) may be added to reduce induction system deposits. *See* PETROLEUM PRODUCTS. [A.C.N.]

Gasterosteiformes

An order of actinopterygian fishes which includes the groups Solenichthyes, Thoracostei, Hemibranchii, Lophobranchii, and Syngnathiformes of recent classifications. The relationship seems to be with the Cyprinodontiformes. Common characters include a swim bladder without a duct; the pelvic fin, if present, with or without a spine and abdominal to subthoracic; a girdle which is not in contact with the cleithrum; articulation of the quadrate with the lower jaw in advance of the orbit; fin spines present or absent; and branchiostegals which are of basic perciform plan or reduced therefrom. The snout is commonly produced, the mouth opening at the end of a long tube.



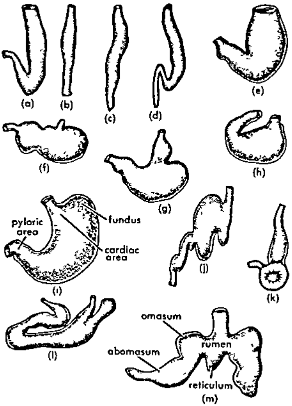
Trumpetfish, Aulostomus maculatus; length to 1 ft. (After D. S. Jordan and B. W. Evermann, The Fishes of North and Middle America, U.S. Natl. Museum Bull. 47, 1900)

Members of this order, commonly known as sticklebacks, tube-snouts, snipefishes, pipefishes, and seahorses, may be classified in 2 suborders, 8 families, about 50 genera, and nearly 200 species. Most are inhabitants of warm seas, but some live in fresh waters of the far north. The group dates from the Eocene. *See* ACTINOPTERYGII. [R.M.B.]

Gastrointestinal tract

That portion of the digestive system which includes the stomach and intestine.

Stomach. The stomach is a muscular, pouchlike organ that is variously modified morphologically in the vertebrates (see illustration). It is absent in the cyclostomes, holocephalans, lungfishes, and a number of teleosts in which the loss may be secor



Types of stomachs found in vertebrates. (*a*) Dogfish. (*b*) Urodele. (*c*) Snake. (*d*) Lizard. (*e*) Toad. (*f*) Hamster. (*g*) Beaver. (*h*) Otter. (*i*) Man. (*j*) Seal. (*k*) Bird. (*l*) Vampire bat. (*m*) Ruminant. (*From H. E. Walter and L. P. Sayles, Biology of the Vertebrates, 3d ed., Macmillan, 1949*)

ary. Three regions may be recognized in the verte-

some fish to the four-chambered stomach of ruminants of which the abomasum constitutes the true stomach. The functions of the stomach include storage of food, mechanical treatment of the food, and initiation of protein digestion. The stomach consists of four layers: an inner mucosa, the submucosa of loose connective tissue, the muscularis, and serosa or peritoneum. Four types of lining epithelium (mucosa) of the stomach may be distinguished: esophageal which is nonglandular and stratified, cardiac which contains tubular glands, fundic with tubular glands that produce digestive ferments (the chief cells) and hydrochloric acid (the parietal cells), and pyloric which contains branched tubular glands. The distribution of this lining epithelium is not necessarily associated with the morphological divisions of the stomach. The muscular layers are usually distributed evenly except in graminivorous birds in which the distal stomach is modified into a heavy muscular gizzard. The epithelial wall of the stomach is deeply folded into ridges (rugae). *See* STOMACH.

Intestine. Two main divisions of the intestine are recognized, the small and large intestine. The small intestine of higher vertebrates is arbitrarily divided into three regions: the duodenum, jejunum, and ileum. Digestion is completed and absorption occurs in the small intestine. Differentiation into the small and large intestine first occurs in the amphibians; the amphibians are also the first to show the presence of villi (minute, fingerlike projections of the mucosa) which increase the absorptive area for digested foods. Internal modifications for absorptive purposes also occur in the cyclostomes where an infolding of the intestine (typhlosole) appears and becomes modified into the spiral valve in elasmobranchs, dipnoans, and certain ganoids. Most ganoids, however, and the teleosts possess a pyloric cecum. The large intestine functions in the disposal of solid waste material and for the resorption of water. This structure is generally a straight, short tube in most vertebrates except the mammals. In many vertebrates a diverticulum may occur at the junction of the small and large intestines. This is the colic cecum and first occurs in the reptiles; however, it is lacking in the crocodilians. Certain birds such as woodpeckers and parrots, and a few mammals also lack a cecum. The distal end of the human cecum is vestigial and prolonged into the vermiform appendix. The large intestine of mammals may be differentiated into ascending, transverse, and descending colon which terminates in the rectum. *See* DIGESTIVE SYSTEM. [C.B.C.]

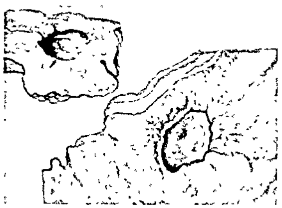
Gastrointestinal tract disorders

Disorders of the esophagus, stomach, and intestine are of frequent occurrence in man and animals. They are of diverse etiology, and the causes of several of the more important ones have not yet been established.

Tumors. Malignant tumors are uncommon in the gastrointestinal tract of animals, but in man more than one-half of all fatal malignant tumors arise in this site. They are most common in the stomach, colon, and esophagus, in that order, and rare in the small intestine. They may be derived from any

wall or projection of the neoplasm into the lumen may cause obstruction, hemorrhage, or perforation. Malignant tumors may also cause death by growth into adjacent tissues or by distant spread (metastases) to the liver, lungs, and brain. Benign tumors of the intestinal tract are composed of proliferating epithelium, lymphoid tissue, fat, or smooth muscle. They may cause obstruction or hemorrhage, and in man, malignant transformation is always a potential hazard. *See* NEOPLASIA.

Peptic ulcers. Ulcers of those parts of the gastrointestinal tract bathed by gastric secretions are known as peptic ulcers. They are, therefore, found primarily in the stomach, duodenum, and, less often, the lower part of the esophagus. Peptic ulcers occur more often in men than in women, especially ulcers of the duodenum. There is a strong nervous or psychological factor in the causation of peptic



Peptic ulcer of the stomach. In the view on the right, a blood vessel extends above the floor of the ulcer crater. Erosion of this vessel caused a serious hemorrhage into the stomach. At the upper left the ulcer is shown in cross section; the crater extends deep into the wall, and there is fibrous scarring at the base.

ulcers and the disease is associated with highly civilized populations and with the more tense and aggressive personality. The ulcers result from digestion of the wall by the action of hydrochloric acid secreted in excessive quantities in these susceptible individuals. The hypersecretion of gastric juice is primarily of nervous origin in those persons with duodenal ulcers, whereas in those with gastric ulcers it is usually on a hormonal basis. In man, the causative significance of local injury is debatable, but in the only two animals known to develop such ulcers spontaneously, physical trauma is very important. Seals ingesting sharp stones with their food develop gastric ulcers, as do weanling calves soon after eating rough fodder.

The symptoms in man are usually characteristic, with gnawing pain which occurs in the upper abdomen 1–4 hours after eating and which is relieved by the ingestion of food. In appearance, peptic ulcers are round craters measuring ¼–1 in. in diameter and extending deep into the wall. Their chronic nature, with partial healing, is manifested by the presence of newly formed connective tissue and blood vessels at the base. With complete healing a tiny stellate scar may be the only residuum. Aside from the distressing symptoms, peptic ulcers are of major significance because of their frequent complications: erosion of a blood vessel may result in hemorrhage; deep extension may cause perforation and peritonitis; edema of adjacent tissues may obstruct the duodenal orifice; and some ulcers may give rise to malignant tumors. Fortunately treatment and cure of peptic ulcers is becoming increasingly effective as the various etiological factors are better understood.

Appendicitis. Appendicitis, or inflammation of the appendix, is the most common surgical emergency. The major cause is thought to be obstruction of the lumen by a small firm mass of fecal material. Normal secretions accumulate in the blind end beyond the obstruction, distending the lumen, compressing blood vessels, and enabling bacteria to invade the wall. Irritation of the outer, peritoneal surface is responsible for abdominal pain. The wall may become sufficiently weakened so that perforation occurs, leading to the development of an abscess in the adjacent tissues, or to generalized peritonitis.

Colitis and regional enteritis. Chronic ulcerative colitis is a disease of unknown etiology, affecting primarily young adults. It is characterized by recurrent episodes of diarrhea, and by progressive involvement of more and more of the large intestine with inflammation, ulceration, and scarring. Cancer of the bowel occurs unduly often in these individuals. A somewhat similar disease of unknown etiology affects segments of the small intestine and is called regional enteritis, or regional ileitis. The involved parts of the bowel become thick-walled and rigid. Ulceration and perforation commonly lead to fistulous communications with other parts of the intestine or with the skin.

Cardiospasm. Achalasia, or cardiospasm, is an obscure condition in which the passage of food through the distal esophagus is impaired. Retention of food leads to remarkable dilation of the entire esophagus. Children sometimes manifest a similar inability to propel the fecal stream through the

stomach, may result from ingestion of many injurious materials, but strong acids are the greatest offenders. Chronic gastritis, however, is a poorly understood condition, eventuating in atrophy of the mucosa and consequent deficiency of gastric secretion. Such atrophy is commonly associated with pernicious anemia and with cancer of the stomach.

Infections. Infections involve primarily the bowel, rather than the esophagus or stomach. In man, the etiological agents include several viruses causing enteritis; bacteria causing typhoid fever, dysentery, and cholera; and parasites ranging from single-cell amebas to large tapeworms. Intestinal parasites are particularly prevalent in birds and animals. Improved sanitation, antibiotics, and other new therapeutic agents have greatly diminished the menace of intestinal infections in man, but they are still responsible for great suffering in many parts of the world. For further discussion, consult the specific infection or infectious agent under its proper name.

Anomalies. Congenital anomalies may occur in any part of the intestinal tract. Particularly important is the failure of the anus or a part of the esophagus or intestine to develop normally. Early surgical treatment is essential if the infant is to survive. Abnormal outpouchings (diverticula) may be either congenital or acquired. Meckel's diverticulum is a common anomaly of the small intestine and is susceptible to acute inflammation. Acquired diverticula occur commonly in the distal part of the colon where they are usually small but numer recurrent inflammation in these divertic

lead to scarring and narrowing of this part of the bowel.

Chemical and mechanical injuries. The esophagus and stomach are primarily affected by such injuries. Particularly tragic is corrosive injury of the esophagus resulting from accidental or suicidal ingestion of strong alkalies, such as lye. Even if perforation is avoided, narrowing (stenosis) invariably accompanies healing. In animals and small children perforation of some part of the intestinal tract may be produced by sharp hard materials ingested with or as food. In animals, such foreign bodies are common causes of intestinal obstruction, but in man obstruction is more often the result of tumors, hernias, fibrous bands (peritoneal adhesions), volvulus (twisting of the bowel or its mesentery), or intussception (telescoping of the bowel into itself). Whatever the cause, intestinal obstruction, with its intermittent cramping pain, is invariably fatal if prolonged. *See* DIGESTIVE SYSTEM. [M.R.H.]

Bibliography: S. L. Robbins, *Textbook of Pathology*, 1957; H. A. Smith and T. C. Jones, *Veterinary Pathology*, 1957.

Gastrolith

Any of the pebbles swallowed by animals and retained for a time in the gizzard or stomach where they serve to grind up the food and in so doing become rounded and highly polished. Birds generally use such pebbles, as do some living reptiles, notably the crocodile and certain lizards. Some of the Meso- [illegible] bedded in formations such as the Niobrara chalk, which is otherwise free of gravel. In some instances as many as a half bushel of such stones have been [illegible] and widespread, however, as to indicate that they are due to natural weathering. No satisfactory criteria are known for distinguishing them from gastroliths. [C.O D.]

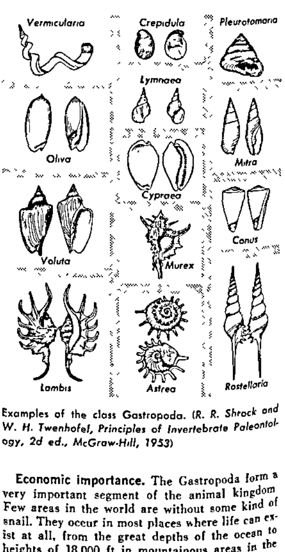
Gastropoda

The largest of the six classes in the phylum Mollusca. Gastropoda contains all of the snails which occur on land and in fresh and salt water. It is divided into three subclasses, the Prosobranchia, Opisthobranchia, and Pulmonata.

Prosobranchia. This subclass is composed mainly of marine snails with a few families found in fresh water and on the land. Torsion has taken place; that is, the visceral hump has become twisted 180°, and the visceral nerve commissures form a figure 8. The right member of each of the paired organs, such as the gills, has been lost, except for a few groups in which the right member of the paired organs has been retained. The gill and anus are in the anterior portion of the body and the gill is in front of the heart. A shell is usually present and it may be caplike, but usually it is coiled to form a multispiral shell. The sexes are separate in the Prosobranchia.

Opisthobranchia. The subclass Opisthobranchia is composed of marine snails with a very few species found in fresh water. In this group, detorsion has occurred, the nerve commissures becoming untwisted. Of the internal paired organs, only those on the left side remain and the gill is behind the heart. Many groups in this subclass are shell-less; in others the shell is small and in a few groups it has become internal. The Opisthobranchia are almost exclusively hermaphroditic.

Pulmonata. The subclass Pulmonata is composed mostly of land and fresh-water groups. Here the gill is replaced by a pulmonary sac, a modification of the mantle cavity. The nerve commissures have become consolidated around the esophagus. All species are hermaphroditic.



Examples of the class Gastropoda. (*R. R. Shrock and W. H. Twenhofel, Principles of Invertebrate Paleontology, 2d ed., McGraw-Hill, 1953*)

Economic importance. The Gastropoda form a very important segment of the animal kingdom Few areas in the world are without some kind of snail. They occur in most places where life can exist at all, from the great depths of the ocean to heights of 18,000 ft in mountainous areas in the tropics. They are important economically because many of the higher forms are predators, feeding

mainly on other mollusks. A limited number, for example, the oyster drills, do considerable damage to oysters and clams in regions where these mollusks are important food for man. A few species of land and marine snails are also important as food for man, particularly in the countries of southern Europe and northern Africa. Many snails are carriers of blood flukes, trematode worms which are responsible for the disease known as schistosomiasis. The snail is an intermediate host in the life cycle of the trematode. Eliminating the particular species of snail carrier in any one region where the disease occurs is still the only practical way of controlling the disease. A family of marine snails, the Conidae, has been responsible for the deaths of several people. They possess a poison apparatus and when carelessly handled they can inflict a poisonous bite which can result in death. One of the top shells, *Trochus niloticus*, is an important source of pearl shell for the button industry in Japan. This mollusk is found mainly in the western Pacific and many hundreds of tons are collected each year for this industry. Abalone shells in the genus *Haliotis* are used in jewelry and the flesh is canned and sold for food. *See* SCHISTOSOMIASIS.

Ecology. Gastropods occupy a large number of specialized habitats. Among the land snails there are many tropical genera which live in trees, feeding on the lichens which grow on the bark and leaves. These tree snails are generally highly colored and are among the most beautiful of all gastropods. Other land snails live only on exposed limestone, and, like the tree snails, feed on lichens. There are many forms that live in desert areas, usually remaining in the soil during dry periods and emerging during any rainy period to feed voraciously on the scant vegetation. Others in desert areas remain attached to the vegetation during dry spells, cementing themselves by means of a limy mucus. Most land snails, however, are found in forested areas, live on the ground, and mainly feed on decaying vegetation at night when it is cool and moist.

Fresh-water snails are found in nearly all types of habitat, from the quiet water of ditches and slow-flowing streams to rapid rivers and even the deep water of large lakes Fresh-water pulmonates are most abundant in clear water in lakes and streams. Many prosobranchs can live in very muddy water. Most fresh-water snails feed on algae.

Marine snails are to be found from well above the strand line to the great ocean depths. Certain groups live only on rocks, others on mud and sand, in and among coral, and on algae. Some feed on algae and a limited few on sea grasses; others are predators, feeding on other mollusks, worms, and echinoderms; and many are scavengers. A few, especially in the genus *Conus*, feed on small fish. A limited few are parasitic and live mainly on echinoderms and on other mollusks. *See* GASTROPODA FOSSILS; MOLLUSCA; *see also* OPISTHOBRANCHIA; PROSOBRANCHIA; PULMONATA. [W.K.E.]

Gastropoda fossils

Fossil gastropods have a long geologic history, being common throughout the Paleozoic and increasingly abundant in the Mesozoic and Cenozoic. All three subclasses, Prosobranchia, Opisthobranchia, and Pulmonata, are known in the fossil record; many superfamilies, particularly prosobranchs, are extinct. Average duration of a genus has been estimated to be from 30,000,000 to 90,000,000 years.

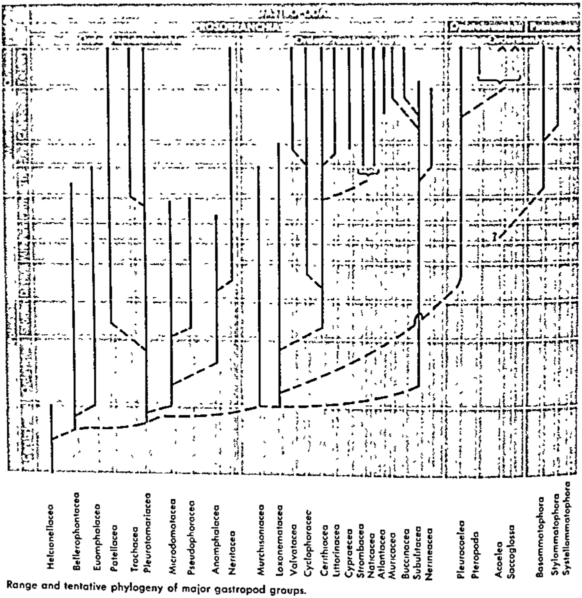
Stratigraphic markers. Marine gastropods are important stratigraphic indicators in Cenozoic strata, and locally are abundant in Cretaceous rocks. They are less common and less useful in the Jurassic and Triassic. Although individual genera have stratigraphic utility within the Paleozoic, it is only in the Ordovician that they are significant for correlation.

In North America, Late Cenozoic nonmarine gastropods are becoming increasingly important for correlation and ecological interpretation of terrestrial deposits. Marine Cenozoic and later Mesozoic gastropods are interpreted as having lived under essentially the same ecologic conditions as related Recent gastropods. The possible ecology of Paleozoic and earlier Mesozoic gastropods has not been thoroughly investigated. In the Ordovician and Silurian, gastropods are most common in limestones. From the Silurian onward, marine forms are more common in shales and some sandstones. Most occurrences of abundant fossil gastropods probably can be interpreted as indicating shallow water conditions.

Evolutionary trends. Prosobranch evolution is best interpreted in terms of gill structure. Primitive Cambrian Helcionellacea are small symmetrical shells with a straight apertural lip. Possibly they had paired gills. Bellerophontacea developed an apertural emargination. This channeled deoxygenated water out of the mantle cavity so that it did not cross the inferred paired gills. Throughout the Paleozoic there were various minor deviations in water circulation which are evident in the modified shape of both the aperture and the emargination.

The asymmetrical shell was a second major development. Primitive Pleurotomariacea are characterized by a sinuslike emargination. In more advanced forms the emargination is slitlike. Unusual apertural shapes parallel those of the Bellerophontacea. These two superfamilies are dominant in the Paleozoic. Living Pleurotomariacea are rare deep-water forms or are modified to a cap-shaped shell for rock clinging.

Loss of one gill, removing the necessity for an apertural emargination, was the next step; probably this happened independently several times. Many Paleozoic forms, thought to have had one gill, superficially resemble Recent forms (Anomphalacea, Pseudophoracea, Microdomatacea). Neritacea and Trochacea, important Recent superfamilies of primitive gastropods, have a long



Range and tentative phylogeny of major gastropod groups.

logic history. Patellacea developed a cap-shaped shell and were specialized for rock clinging like some modern Pleurotomariacea.

Two other specializations particularly characterized branches of the Paleozoic one-gilled plexus. Platyceratacea adapted to a coprophagous life on crinoid calices and became extinct coincidentally with primitive crinoids. Euomphalacea perfected low-spired shells, a few approaching bilateral symmetry.

A fourth evolutionary step was a change in the nature of the remaining gill. Accompanying this [illegible]

were confined to hard bottoms because of the possibility of fouling their gills with mud. Development of a siphon allowed gastropods to travel below the mud-water interface with the siphon projecting up into the water.

Geological distribution. Siphonate genera first appear in the Ordovician. Murchisoniacea are intermediate in possessing both a basal fold and an apertural emargination. Loxonematacea had both a fold and an apertural sinus. Subulitacea possessed only a fold. All these are persistent throughout the Paleozoic but are not abundant.

Siphonate genera are slightly more common in the upper Paleozoic and increase markedly in the Mesozoic and Cenozoic. Most Paleozoic siphonate prosobranchs are fusiform with a simple siphonal fold. [illegible] apertures [illegible] ture by [illegible] ea) are post-Paleozoic characteristics. The stratigraphically important Nerinacea developed elaborate internal shell thickenings.

Ecological characteristics. Both primitive and more advanced single-gilled forms were able to invade brackish and fresh water habitats (Valvatacea,

Cyclophoracea, Rissoacea, Cerithiacea). Nonmarine gastropods are first known in the Pennsylvanian but may have evolved earlier. The most specialized prosobranchs, Atlantacea, developed a thin shell, secondarily symmetrical, and took up a pelagic life.

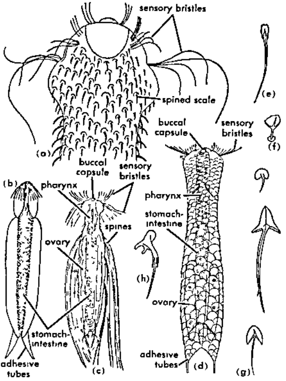
Opisthobranchs can be differentiated from fossil prosobranchs by their characteristic deviated early shell. They are rare in the Devonian (Pleurocoelea) but are more common in the Mesozoic and Cenozoic. Some of the rarity may be spurious, the result of failure to collect smaller fossils and lack of interest in the suborder. An important trend is a gradual return to bilateral symmetry, coincident with a reduction of the shell. Important shell-less Recent forms are not represented in the fossil record. A second trend has been toward parasitism. A third trend, Aplysiacea, is to a pelagic existence with a shell resembling that of the Atlantacea.

The pulmonates are the most abundant and varied of the living gastropods. They may have evolved in a marine habitat, and then spread to fresh water and terrestrial environments at a later time. A few still occupy marine habitats today. The exclusively aquatic Basommatophora are first known with certainty in the Triassic. At that time, the major superfamilies had already differentiated. It is possible that some of the Paleozoic nonmarine forms may eventually be classed as ancestral pulmonates. The more varied Stylommatophora are less well known than the Basommatophora. Their fossil record before the Cretaceous is extremely poor. The shell-less Systellommatophora are not represented as fossils and therefore are of unknown age. [E.L.Y.]

Gastrotricha

A group of microscopic animals considered either to be a class of the phylum Aschelminthes or to constitute a separate phylum. These cuticulate animals are multicellular, pseudocoelomate, unsegmented and somewhat wormlike. Two orders occur, the Macrodasyoidea which are exclusively marine and the Chaetonotoidea which are mainly freshwater forms. Characteristic features of the group are the cilia used mainly in locomotion, a cuticular pattern of plates, scales, and in some species spines. Externally, the body ranges from complete smoothness and transparency in species of *Ichthydium* through a very intricate display of short to long spinous structures and scales which entirely clothes the cellular body.

Reproduction and development. A gastrotrichan egg is elliptical with one surface somewhat flattened. The protective cuticula is often smooth but characteristically is provided with spines, raised protuberances, and roughnesses of variable shape and number. Parthenogenesis prevails in freshwater species, if, indeed, it is not exclusively exhibited. Before depositing eggs, the female is markedly swollen by the fully mature ovum and in lateral view looks like an ant with a prominent abdomen. *See* PARTHENOGENESIS.

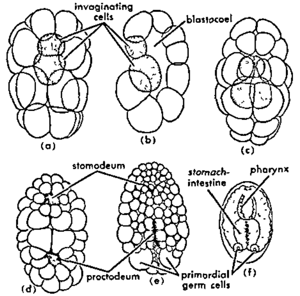


Scalation. (a) Anterior end of *Chaetonotus*, showing sensory bristles and spined scales (*after C. Zelinka, 1889*). (b) *Proichthydium* (*after E. Cordero, 1918*), with a ciliary row across the head. (c) *Stylochaeta*, from life, with bunches of long spines. (d) *Lepidodermella*, from life, clothed with spineless scales. (e, g, h) Types of spined scales (*after C. Zelinka, 1889; and M. Voigt, 1904*). (f) Stalked type of scale. (From L. H. Hyman, *The Invertebrates, vol. 3, McGraw-Hill, 1951*)

Two kinds of eggs have been described from the standpoint of an immediate and a delayed cleavage. The former start dividing as soon as deposition occurs. They hatch from 12 to 20 hours later into a small replica of the adult which attains full size several hours after emerging from the shell.

The slightly larger eggs laid in older cultures, as environmental conditions become unfavorable, [illegible] long as 2 years before resuming divisional activity.

Development is direct. The first few cleavages resemble the pattern followed by Nematoda, but after this all likeness is lost. Blastomeres are almost equal and cleavage is total. Little of the organology can be traced with exactness other than the formation of the digestive tract. Marine species appear to have a larger number of cells than members of the Chaetonotoidea where there are indications of cell constancy within the species. A gastrotrich has a short life history of from 3 to 21 days. This appears to be dependent upon the species and prevailing ecological conditions. Their life span in cultures is as long as 3 weeks and one individual may



Embryology. (a) Beginning gastrulation, two invaginating cells stippled. (b) Sagittal section of a. (c) Later stage, invaginated cells have divided to four. (d) Formation of stomo- and proctodeum. (e) Stomo- and proctodeum completed, primordial germ cells differentiated. (f) Advanced embryo in eggshell. (*Examples of genus Neogossea, after P. Beauchamp, 1929 from L. H. Hyman, The Invertebrates, vol. 3, McGraw-Hill, 1951*)

lay five eggs. The cuticula is not shed during life. *See* CELL CONSTANCY.

Ecology. Too little is known of the exact environmental factors under which gastrotrichs live. They are readily found in standing fresh waters where algae and protozoa abound, most likely close to or within the delicate film of the substrate although some species seem to move about from point to point within a depression slide drop taken from a collection or set up as an artificial culture. Some studies have noted the greatest concentration in *numbers during the months of July and August in* outdoor pools and ponds; however, samples taken from beneath ice in January have yielded specimens and they have been found in their native habitats during all months of the year.

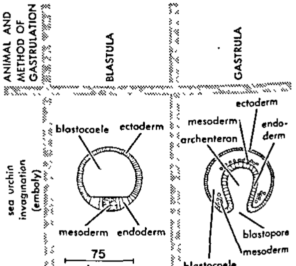
Gastrotrichs possibly evolved from the Turbellaria and show a distinct relationship to the Nematoda. [C.E.P.]

Bibliography: L. H. Hyman, *The Invertebrates*, vol. 3, 1951; R. W. Pennak, *Freshwater Invertebrates*, 1953.

Gastrulation

The formation of the primordial gut, the archenteron, or digestive cavity of an early animal embryo. More generally, and originally, the term gastrulation referred to the process by which the gastrula stage of the embryo is formed. Thus, to nineteenth-century embryologists, gastrulation was the process by which the single-layered blastula, a hollow ball of cells, is converted into the double-layered gastrula (*see* BLASTULATION). The term has now come to have a still more general meaning, namely, the process by which the three germ layers, or primordial tissues of the embryo, are brought into the positions and relations characteristic of the late gastrula stage, with ectoderm (outer skin), mesoderm (middle skin), and endoderm (inner skin) from the outside to the inside. See illustration (*b*), (*d*), and (*f*). The terms epiblast, mesoblast, and hypoblast are also used to denote ectoderm, mesoderm, and endoderm, respectively.

Process of gastrulation. In the blastula stage, before the beginning of the process of gastrulation, the prospective germ layer regions, or groups of cells which normally form the germ layers, can be



Diagrams of blastula and gastrula stages of some animal embryos: (a) sea urchin blastula; (b) sea urchin gastrula; (c) amphibian blastula; (d) amphibian gastrula; (e) amniote blastula; (f) amniote gastrula.

detected, localized, and mapped out by means of various methods. These consist of marking or tagging trial areas on the surface of the blastula with nontoxic dyes or inert colored powders such as carbon black, carmine, and chalk. The marked areas are then followed during gastrulation (*see* FATE MAPS, EMBRYONIC). In vertebrate embryos, other prospective organs or tissues have similarly been mapped on the surface of the blastula. These include the neural plate, future brain and spinal cord; notochord, a medial mesodermal rod around which the vertebral column of the skeleton is later formed; and somites, mesodermal cell aggregates which contribute to formation of the vertebrae, muscles of the back and other structures. Thus, gastrulation involves the movement and rearrangement of the prospective organ-forming or germ-layer-forming regions, or both, of the blastula, into a special pattern which one can recognize as the primary body plan, characteristic of the particular species.

Consequently, gastrulation means, to most modern embryologists, a complex of processes extending over a variable period of time and involving various patterns of coordinated cell movements, called morphogenetic movements. In this, formation of the primitive gut is only one of many events, but establishment of the fundamental body plan is the major result. There is great variety in the modes of gastrulation throughout the animal kingdom, even among rather closely related animals, much of which is referable to differences in the cellular composition or architecture of the blastula and also to differences in egg structure, such as the size of the egg and amount and distribution of yolk. However, many embryologists prefer to consider the various patterns of gastrulation as means to a common end or as variations on a basic theme which, in a general way, seem to relate even animals which appear very different.

A definition and characterization of the blastula stage is prerequisite to a clear understanding of the processes and methods of gastrulation. Some confusion in the understanding of the processes of gastrulation seems to stem from the failure of embryologists to agree upon the nature and significance of the blastula stage. Thus, what some have described as a process of gastrulation, in endoderm formation by delamination, might better be considered as a process of blastula formation.

Methods of gastrulation. Two general but not mutually exclusive methods of gastrulation have been recognized: epiboly and emboly. Epiboly is the growing or extending of one part, such as the upper hemisphere of a spherical blastula, over and around another part, like the lower hemisphere. Emboly is the pushing or growing of one part into another as shown in the illustration (*b*) and (*d*). In many embryos, both types of cell movement may occur; in certain invertebrate embryos, one type may predominate almost to the exclusion of the other. Generally speaking, epiboly tends to be the major, but not the only, method of gastrulation in forms with large, yolky eggs. In some animals, such as certain coelenterates, no pronounced epibolic or embolic movements of cells seem to occur during gastrulation, the gut cavity, or coelenteron, being formed by a process of cavitation, in which spaces arise within a solid mass of endoderm cells and coalesce to form the primitive digestive cavity. In other invertebrates, the gut cavity may form by cavitation, with other processes of gastrulation being accomplished by epiboly and emboly.

Epiboly is a very common type of cell movement, involving a peripheral spreading or expansion of the epiblast over the hypoblast, and, in some embryos, such as amphibians, over the mesoblast as well. Epiboly of the epiblast over the hypoblast apparently does not occur in the blastoderm type blastula of reptiles and birds. Here extraembryonic epiblast and hypoblast spread together at the margin of the blastoderm, over the relatively large yolk, by a process having little or no direct relation to the events of gastrulation, which are restricted to the more central area of the blastoderm.

The cell movements involved in emboly bring about the inward movement of the prospective endoderm and mesoderm cells. Usually, the embolic movements occur only in a fairly restricted or circumscribed area on the surface of the blastula. This area usually indents to form a pore or pit and is termed the blastopore. Three general types
of embolic
vagination.
in, of a por
sulting in the formation of a cavity, the primitive gut or archenteron, bounded by the infolded cells. This is exemplified by the endoderm invagination in the sea urchin. (2) Involution. This is an inrolling or inward rotation of the blastula surface at, or near, the edge or lip of the blastopore which has already been formed by an initial, but somewhat minor invagination movement. (3) Ingression. This is the inward movement of single or small groups of cells, in contrast to a sheet or layer of cells, from the surface into the interior. Examples include the primary mesoderm ingression in the sea urchin, endoderm ingression at the vegetal pole of the amphibian gastrula, and, possibly, some mesoderm ingression through the early blastopore, or primitive streak, of amniotes. Vertebrate gastrulation may involve many, if not all, of the above types of cell movements. *See* GERM LAYERS. [N.T.S.]

Bibliography: L. B. Arey, *Developmental Anatomy*, 6th ed., 1954; E. W. MacBride and J. G. Kerr, *Textbook of Embryology*, vols. 1 and 2, 1914; O. E. Nelsen, *Comparative Embryology of the Vertebrates*, 1953.

Gate circuit

An electronic circuit having one output and one or more inputs, in which the output is a function of the inputs in a prescribed manner and in a controllable time sequence. Gate circuits may be classified as transmission and switching gates.

Transmission gate. The transmission gate is defined as one in which the output waveform is a replica of a selected input during a specific time interval. The control signal, which determines the time interval, is the gating signal; the gate generator is often a multivibrator. A simple transmission gate having only one signal input is shown in Fig. 1. The gating signal is derived from the monostable multivibrator, which is triggered at time t_1 from an external source. The multivibrator generates a gating pulse having a period T. The gate signal is applied to the suppressor grid of the pentode and allows the tube to function as an amplifier during the time T, thereby allowing transmission of the signal waveform v_s. Other signals could be "gated in" at other times by connecting the plates of additional tubes in parallel and applying such signals to the control grids and appropriate gating waveforms to the suppressor grids.

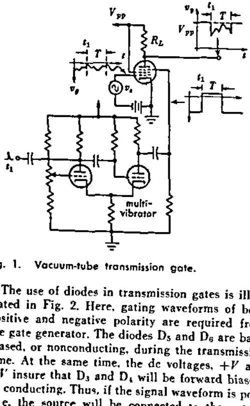


Fig. 1. Vacuum-tube transmission gate.

The use of diodes in transmission gates is illustrated in Fig. 2. Here, gating waveforms of both positive and negative polarity are required from the gate generator. The diodes D_5 and D_6 are back biased, or nonconducting, during the transmission time. At the same time, the dc voltages, $+V$ and $-V$ insure that D_3 and D_4 will be forward biased, or conducting. Thus, if the signal waveform is positive, the source will be connected to the output through the low resistance of D_1 and D_3, and if negative, through D_2 and D_4. During the nontransmission time, the outputs of the gate generator will be connected through D_5 and D_6, and the signal-path diodes will be reverse biased, or nonconducting.

Switching gate. In the switching, or logical, gate an output that is not a replica of the input is registered during a time interval if a particular combination of input signals exists. The output only indicates that inputs exist in a certain sequence or combination and yields no information as to amplitude. Examples are the OR, AND, NOT and INHIBIT circuits which are basic building

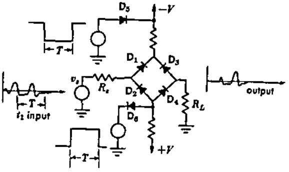


Fig. 2. Diodes in transmission gates.

blocks of digital computers. Such circuits use combinations of multivibrators and diodes in much the same manner as transmission gates, and combinations of such functions are usually referred to as switching circuits. *See* SWITCHING CIRCUIT.

[C.M.C.]

Gauss

A unit of magnetic flux density in the centimeter-gram-second (cgs) electromagnetic system of units.

If the cgs electromagnetic system is used in the defining equation for magnetic induction or flux density B,

$$B = \frac{F}{Il \sin \theta}$$

the force F is in dynes, the current I is in abamperes, and the length l is in centimeters. The unit of B is the dyne/abampere-centimeter. This unit is 10^{-4} newton/amp-m.

In this system, the choice of lines to represent the magnetic induction quantitatively is made by taking the number of lines per square centimeter of a surface perpendicular to B as equal to B in cgs units. A line thus defined is called a maxwell. A maxwell per square centimeter is called a gauss. A gauss is equivalent to a dyne/abampere-centimeter. It follows that 1 weber/m² = 10^4 gausses. *See* ELECTRICAL UNITS; INDUCTION, MAGNETIC; MAXWELL; WEBER.

[K.V.M.]

Gauss' theorem

The assertion, under certain light restrictions, that the volume integral through a volume V of the divergence of a vector function $\mathfrak{F}(x,y,z)$ is equal to the surface integral of the exterior normal component of $\mathfrak{F}$ over the boundary surface S of V, or in symbols $\iiint \nabla \cdot \mathfrak{F}\, dV = \iint \mathfrak{F} \cdot \mathbf{v}\, dS$ with $\mathbf{v}$ the unit exterior normal to S. This theorem is also known as the divergence theorem and as Green's theorem.

A simple illustration is provided by the radius vector $\mathbf{r}$, $(\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{f})$ and the sphere $S(0,a)$ (center at O, radius a). Because $\nabla \cdot \mathbf{r} = 3$ and on S, $\mathbf{v} = \mathbf{r}/a$, $\mathbf{r} \cdot \mathbf{v} = a$, it follows that $\iiint \nabla \cdot \mathbf{r}\, dV = 3(\frac{4}{3}\pi a^3)$ while $\iint \mathbf{r} \cdot \mathbf{v}\, dS = a(4\pi a^2)$.

A common method of proof consists of the conversion of $\iiint \nabla \cdot \mathfrak{F}\, dV$ into a sum I of three double integrals with domains in the coordinate planes followed by the conversion of I to $\iint \mathfrak{F} \cdot \mathbf{v}\, dS$. The

function $\mathfrak{F}$ may be assumed to be of class C' and the volume V to be convex with smooth boundary pieces. The class of admissible volumes may be extended by the introduction of protuberances and cavities of the previous type.

The divergence theorem plays an important role in a variety of subjects such as electricity and magnetism, mechanics of continuous media (including fluid dynamics), heat flow, partial differential equations, and potential theory. *See* CALCULUS OF VECTORS; GREEN'S THEOREM. [H.V.C.]

Gaviiformes

The order of birds containing the loons or divers. There is but a single family, Gaviidae, with four living species nesting in cooler parts of the Northern Hemisphere. The best known and most southerly species is the common loon or great northern diver (*Gavia immer*). Loons are goose-sized, fish-eating, diving birds and are highly adapted for aquatic life. Their legs are placed so far back on the body that they can neither walk well nor gain the air from dry land. The nest is usually placed at the water's edge and only two eggs are laid. The loons thus differ from the superficially similar grebes which usually raise large families. The feet are strongly webbed, the tarsi greatly compressed and bladelike, and the bill heavy and pointed. *See* AVES. [K.C.P.]

Gear

A machine element used to transmit motion between rotating shafts when the center distance of the shafts is not too large. Toothed gears provide a positive drive, maintaining exact velocity ratios between driving and driven shafts, a factor that may be lacking in the case of friction gearing which is subject to slippage (*see* ROLLING CONTACT) While the motion transmitted by mating gears is kinematically equivalent to that of rolling surfaces identical with the gear pitch surfaces, the action of one gear tooth on another is generally a combination of rolling and sliding motion. When the distance between shafts is large, other methods of transmission are used. *See* BELT DRIVE; CHAIN DRIVE; ROPE DRIVE.

The application of gears for power transmission between shafts falls into three general categories: those with parallel shafts, those for shafts with intersecting axes, and those whose shafts are neither parallel nor intersecting Examples of these are, respectively, spur gears, bevel gears, and worm gears.

Principal features. Figures 1 and 2 illustrate the principal features of toothed gears. Such terms as pitch circle, addendum circle, and root circle, being geometrical, are defined by the diagrams.

Definitions for other commonly used terms for describing gears are addendum, the radial distance between the pitch circle and the addendum circle; dedendum, the radial distance between the pitch circle and the root circle; face, the tooth surface outside the pitch circle; flank, the tooth surface between the pitch circle and the root circle; clearance, the amount by which the dedendum exceeds the addendum of the mating gear; whole depth, addendum plus dedendum; and working depth, whole depth minus clearance.

Terms for describing toothed gears are pitch point, the point of tangency of the pitch circles of two mating gears; common tangent, the line tangent to the pitch circles at the pitch point; point of contact, any point at which two tooth profiles touch each other; line of action, the line passing through the point of contact normal to a pair of teeth in mesh; pressure angle, the angle between the line of action and the common tangent (this is constant for involute gear teeth but varies with the point of contact for cycloidal teeth); path of contact, the path traced by the successive points of contact of any two teeth; arc of action, the arc of the pitch circle through which a tooth profile moves from the beginning to the end of contact with a mating profile; arc of approach, the arc of the pitch circle through which the tooth profile moves from first contact until contact at the pitch point; the angle of approach is that subtended by this arc; arc of recession, similar to arc of approach, except from the pitch point to the last contact; and contact ratio, the ratio of the arc of action to the

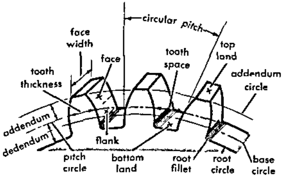


Fig. 1 Principal features of gear teeth.

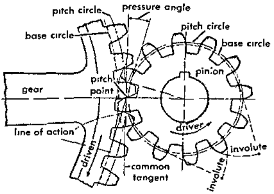


Fig. 2. Characteristics of spur gears. (From C. Ham, E. J. Crane and W. L. Rogers, Mechanics of Machinery, 4th ed., McGraw-Hill, 1948)

circular pitch. The contact ratio may be thought of as the average number of teeth in contact.

Gear tooth sizes are designated by diametral pitch, which is the number of teeth per inch of diameter of the pitch circle. Pitch circle is, in turn, the circle whose periphery is the pitch surface, or surface of an imaginary cylinder that would transmit by rolling contact the same motion as the toothed gear. A gear with a 20-in. pitch diameter and teeth having a diametral pitch of 2 has 40 teeth. Diametral pitch P times circular pitch p (Fig. 1) equals π.

The smaller of two gears in mesh is the pinion. It has the fewer teeth and is driven in a speed reduction unit. The minimum number of teeth an involute pinion can have and still run without interference is fixed by the tooth system. Smaller pinions are possible if the tooth flanks are undercut.

Backlash. The amount by which the tooth space of a gear exceeds the tooth thickness of the mating gear at the pitch circle is the backlash. It can be determined in the plane of rotation or, for helical gears, in the plane normal to the tooth face.

If mating gears were to have zero backlash, gears and mountings would need to be dimensionally perfect. To retain zero backlash with varying operation conditions, all parts need exactly the same thermal expansion characteristic. Because of the difficulty of meeting these requirements and for lubrication, freedom—backlash—is provided between gear teeth. The usual practice is to reduce the tooth thickness on each gear by an amount equal to half the desired backlash. However, in the case of a gear and a small ... [illegible]

... in the pressure angle, the action of involute gear teeth is not affected by backlash or center distance adjustment.

In the case of precision gearing for control systems and similar applications, backlash results in a nonlinear relation between input and output. Several methods of reducing backlash are in use. One method is to place two identical spur gears on the same shaft, one fixed to the shaft, one free. The loose gear is attached to the fixed one by springs, which keep th... [illegible]

... ment along the shaft to eliminate excessive backlash.

Gear action. A principal function of gears is to change the speed of rotation. This action is described by the velocity ratio of the gears in mesh. Ratio VR is the number of revolutions N_1 of the driving gear divided by the number of revolutions N_2 of the driven gear in the same time interval. For gears with teeth T_1 and T_2 respectively

$$VR = N_1/N_2 = T_2/T_1$$

When two curved surfaces, such as the mating surfaces of two gear teeth, are in driving contact there is a definite velocity ratio between the bodies. The angular velocities of the two bodies are inversely proportional to the segments into which their line of centers is divided by a line passing through their point of contact and normal to their surfaces at this point. Thus a constant angular velocity ratio between bodies in driving contact demands that the common normal to the profiles at the point of contact cut the center line at a fixed point, the pitch point. This latter statement is frequently referred to as the fundamental law of gear tooth action. Pure rolling contact between gear teeth occurs only when they are in contact at the pitch point. At all other positions the teeth have some sliding with the maximum sliding at the first and last instants of contact. While there are a number of tooth forms that will satisfy the fundamental law, only two of them have been used to any great extent. The *cycloidal tooth* predominated until the late 1800s but has been replaced to a great extent by the involute gear tooth. Cycloidal teeth are still found in instruments, watches, clocks, and, occasionally ... [illegible]

... cycloidal gears to be interchangeable, all teeth must be generated with the same describing circle and they must have the same pitch, addendum, and dedendum. For involute gears to be interchangeable they must have the same pitch, pressure angle, addendum, and dedendum.

Involute gear teeth. An involute tooth is laid out ... [illegible]

... circle of the involute. The proportions of the tooth are fixed by the gear tooth system and the diametral pitch. The involute curve establishes the tooth profile outward from the base circle. From the base circle inward, the tooth flank ordinarily follows a radial line and is faired into the bottom land

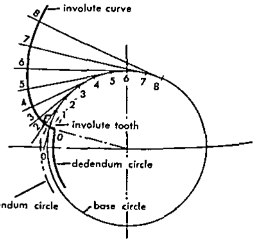


Fig. 3. Method of constructing shape of involute tooth.

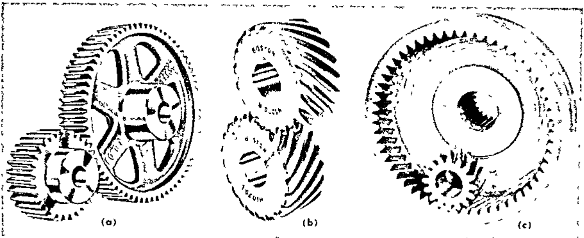


Fig. 4. Spur gears. (a) External spur gear and pinion. (b) Single helical gear (*Boston Gear Works*). (c) Helical internal gear and pinion. (*The Fellows Gear Shaper Co.*)

with a fillet. The basic rack form of the involute tooth has straight sides, an important property from the manufacturing standpoint.

Gear teeth may interfere with each other, especially where pinions have a relatively small number of teeth. Normal contact of involute gear teeth is along the tangent to the base circles of the two gears in mesh (Fig. 2). When contact takes place off this line, interference occurs; that is, the tip of one tooth digs into the flank of another. There are several ways to eliminate interference: (1) reduce addendum of the gear, (2) increase pressure angle, (3) increase backlash by increasing center distance between gears, (4) undercut flank of the pinion, and (5) relieve or modify face of the gear tooth

Spur gears. In the truest sense, spur gears are only those that transmit power between parallel shafts and have straight teeth parallel to the gear axis (Fig. 4*a*). It is common practice, however, to group helical gears that have parallel shafts under the heading of spur gears (Fig. 4*b*). The pitch surfaces of gears with parallel axes are rolling cylinders, and the motion these gears transmit is kinematically equivalent to that of the rolling pitch cylinders.

Spur gears are classified as external, internal, and rack and pinion. External spur gears, the most common, have teeth which point outward from the center of the gear. Internal or annular gears have teeth pointing inward toward the gear axis (Fig. 4*c*). A rack may be considered as a gear having an infinite pitch circle radius. Thus its pitch surface is a plane. A rack and pinion running together transform rectilinear motion into rotary motion, or vice versa.

To standardize the manufacture of gears, four types of spur gear teeth are commonly used, two with pressure angle of 14½° and two with pressure angle of 20°.

An internal gear has the positions of the addendum and dedendum reversed from those of an external gear. This results in a different tooth and less slippage than with an equivalent external spur gear. For a given tooth ratio, the arc of action of an internal gear is slightly greater than that of an external gear of the same size and the tooth is stronger. The nature of an internal gear makes it especially suited to closer center distances than could be used with an external gear of the same size. When it is necessary to maintain the same sense of rotation for two parallel shafts, the internal gear is especially desirable since it eliminates the need for an idler gear. These conditions make the internal gear highly adaptable to epicyclic and planetary gear trains.

Noncircular gears are used to obtain velocity ratios that vary in a precise manner. Elliptical gears are an example of noncircular gears. They provide a convenient method of obtaining a quick-return for machines that do most of their work during only a portion of the drive shaft revolution. Noncircular gears are used in computing mechanisms and other devices where a prescribed varying output function is to be obtained using a linear input.

Helical gears. Gears running on parallel axes and with teeth twisted oblique to the gear axis are essentially spur gears. Because of the twist, contact is progressive across the tooth surface, starting at one edge and proceeding across the face of the tooth. The action results in reduced impact and quieter operation, particularly at high speed. Double helical gears with helixes of opposite hand are called herringbone gears. They are especially suited for high-speed operation and eliminate the axial thrust produced by single helical gears. Dimensions of a helical gear are determined on both the plane of rotation and on the plane perpendicular to the helix angle of the tooth. Thus, a helical gear has a circular pitch measured on the plane of rotation and a normal circular pitch measured on the normal plane.

Crossed helical gears. Where shafts cross obliquely, motion is transmitted by crossed helical gears (Fig. 5). The teeth are helical but differ

the teeth of worm gears in that no one tooth (thread) makes a complete turn on the pitch circle. Pitch surfaces of crossed helical gears are cylindrical as with spur gears. However, with crossed shafts, the oblique teeth have point contact rather than the line contact that occurs with parallel shafts. Analysis of the gears is based on equal components of the pitchpoint velocity on each mating gear along the common normal *N-N*. Sliding occurs in the direction *T-T* of the tooth elements. As with spur gears, the number of revolutions per unit time are inversely proportional to the numbers of teeth. Velocity ratio *VR* may also be expressed as

$$VR = \frac{N_1}{N_2} = \frac{D_1 \cos \psi_1}{D_2 \cos \psi_2}$$

where *D* is pitch diameter and ψ helix angle. Helix angle is between the shaft axis and a line tangent to the tooth through the pitch point.

Helical gears are referred to as right- or left-hand in the same manner as screw threads, a right-hand gear being one on which the teeth twist clockwise as they recede from an observer looking along the axis.

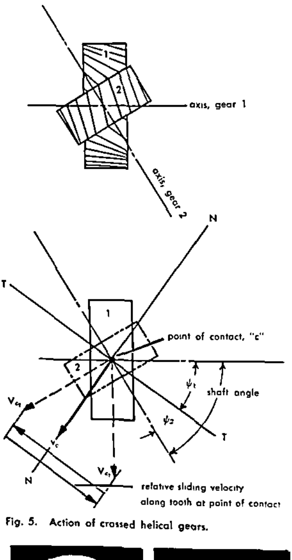


Fig. 5. Action of crossed helical gears.

Bevel gears. Where shafts intersect, bevel gears transmit the motion. Such gears may be used only to change the shaft axis direction or to change speed as well as direction. Two bevel gears with equal numbers of teeth and running together with their shaft axes intersecting at 90° are called miter gears. Several forms of bevel gears are in use, including straight-tooth, spiral, and skewed bevel gears (Fig. 6).

External bevel gears have pitch angles less than 90° (Fig. 7*a*). Internal bevel gears have pitch angles greater than 90°, hence their pitch cones are inverted (Fig. 7*b*). A crown gear is one having a pitch angle of 90° (Fig. 7*c*). Thus its pitch surface is a plane, and the crown gear corresponds in this respect to a rack in spur gearing.

Straight bevel gears. The simplest form of bevel gear has straight teeth which, if extended inward, would come together at the intersection of the shaft axes. This point is also the apex of the rolling cone, which forms the pitch surface of the gear. Much of the terminology applied to bevel gears is the same as that used for spur gears. Additional terms used in reference to bevel gears are given in Fig. 8. Diametral pitch of a bevel gear is not constant across the full width of the tooth. The diametral pitch at the pitch diameter is used in fixing tooth proportions. The formative number of teeth in a bevel gear is the number of teeth that would be on a spur gear whose pitch radius equalled the back cone distance of the bevel gear.

The speeds of the shafts of bevel gears (velocity ratio) are inversely proportional to the numbers of teeth on the gears or to the sines of the pitch

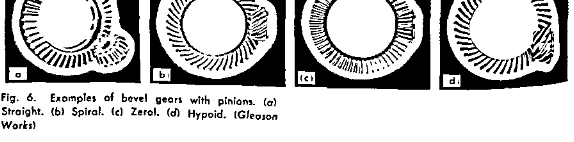


Fig. 6. Examples of bevel gears with pinions. (*a*) Straight. (*b*) Spiral. (*c*) Zerol. (*d*) Hypoid. (*Gleason Works*)

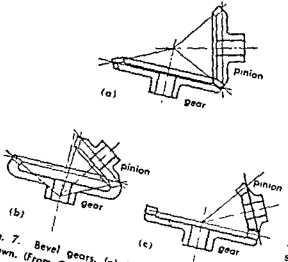


Fig. 7. Bevel gears. (a) External. (b) Internal (c) Crown. (From G. L. Guillet and A. H. Church, Kinematics of Machines, 5th ed., Wiley, 1950)

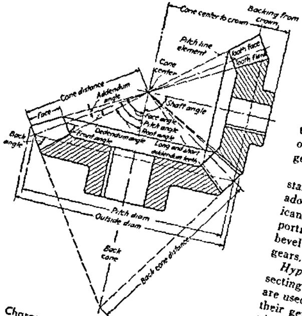


Fig. 8. Characteristics of bevel gears (From C. W. Ham, E. J. Crane and W. L. Rogers, Mechanics of Machinery, 4th ed., McGraw-Hill, 1948)

angles, but not to the formative number of teeth. Use of straight bevel gears is limited to low-speed operations, ordinarily below 1000 surface feet per minute or, in the case of small gears, 1000 rpm.

Since each point on a straight tooth bevel gear remains a fixed distance from the pitch cone apex, there is no sliding along the tooth as it engages. Contact across the full tooth face occurs instantaneously—as with spur gears—as the teeth come into mesh.

Spiral bevel gears. To provide a gradual engagement, as contrasted to the full line engagement of straight bevel gears, the teeth of spiral bevel gears are curved and oblique. Theoretically the curve is a spiral but, to facilitate manufacture, the curve is actually a circular arc which, within the tooth face width, closely approximates a spiral. This tooth inclination brings more teeth in contact at any one time than with an equivalent straight-tooth bevel gear. The result is smoother and quieter operation, particularly at high speeds, and greater load-carrying ability than with straight bevel gears of the same size.

Spiral bevel gears are used in sewing machines, motion picture equipment, machine tools and other applications where quiet, smooth operation is essential. They should, in general, be mounted on antifriction bearings because of the axial thrust due to the oblique teeth. In the past they have been used extensively in the rear axle drives of automobiles but are now being replaced by hypoid gears.

Zerol bevel gears. A special form of bevel gear has curved teeth with a zero degree spiral angle. Thus the teeth are not oblique as is the case with spiral bevel gears. Rather, the teeth lie in the same general direction as those of an equivalent straight-tooth bevel gear, and so the gears are usually used in the same types of drives as the straight-tooth gear. As with straight bevels, they produce no axial thrust and, therefore, may be used without thrust bearings. The fact that they may be produced on the same equipment as spiral, bevel, and hypoid gears makes them economically desirable.

Tooth proportions for bevel gears follow the standards established by the Gleason Works and adopted as the recommended standard of the American Gear Manufacturers' Association. Tooth proportions are a function of the velocity ratio. Thus, bevel gears are not as interchangeable as spur gears.

Hypoid gears. To connect nonparallel, nonintersecting shafts, usually at right angles, hypoid gears are used. They are similar to spiral bevel gears in their general appearance. The axis of the hypoid pinion may be offset above or below the axis of the gear. The shape of the tooth is similar to that of the spiral bevel gear and gives progressive contact across the tooth. In operation these gears run even more smoothly and quietly than spiral bevel gears. To maintain line contact of the teeth, with the offset shaft, the pitch surface of the hypoid gear is a hyperboloid of revolution rather than a cone as in bevel gears.

One of the first uses of hypoid gears was in the rear-axle drive of Packard automobiles. The operating smoothness of hypoid gears, along with the lower body lines made possible by the offset pinion shaft, has made them extremely popular for automotive use. Industrial applications of the hypoid gear also take advantage of the pinion offset, which allows the mounting of any number of pinions on a single continuous shaft, a feat not possible

bevel gears. The shaft arrangement of the hypoid gear and pinion enables bearings to be placed on both sides of the gear and of the pinion. The offset axis results in a larger and, consequently, stronger hypoid pinion tooth than on an equivalent straight-tooth or spiral bevel gear. Expressed in terms of pitch diameter, a hypoid pinion has fewer teeth than a spiral bevel pinion of the same pitch diameter. It is possible to use hypoid pinions having 7, 8, or 9 teeth in contrast to a minimum of 12, 13, or 14 teeth—depending on the velocity ratio—on a spiral bevel pinion. Lubricants must withstand the higher loading and the sliding that occurs along the teeth of hypoid gears.

Hypoid gears are suitable for large velocity reductions; reduction ratios of 60:1 and higher are entirely feasible. In general, shaft offset should not exceed 40% of the equivalent bevel gear back cone distance and, when the loading is very heavy as in truck and tractor drives, the offset should be nearer 20% of this distance. Direction of the offset, above or below center, must be specified for any given installation. The gears should, in general, be mounted on antifriction bearings in an oiltight case. Thrust bearings must be provided. Due to the sliding tooth action, the efficiency of hypoid gears is somewhat less than that of equivalent bevel gears.

Worm gears. A chief way to connect nonparallel, nonintersecting shafts that are at right angles is through a worm gear. The worm, ordinarily the driver, is similar to a crossed helical gear except that it has at least one complete tooth (thread) around the pitch surface. The mating gear is the worm wheel or worm gear. Worm gearing is generally used to obtain large velocity reductions with the worm as the driver and the worm wheel as the driven gear, although occasional applications, for example cream separators, have the worm wheel as the driver. The pitch surfaces of straight worms are cylinders and the involute teeth have point contact. Because the appearance of the worm is similar to that of a screw, the teeth are called threads.

The pitch of the worm is the axial distance from any point on one tooth to the corresponding point on the next tooth (Fig. 9). This must equal the circular pitch of the mating worm wheel. Lead is the axial distance the worm helix advances in one complete revolution around the pitch surface. A single thread worm has the pitch and lead equal; one revolution of such a worm will, for shafts at right angles, advance the worm wheel $1/N$ revolutions if N is the number of teeth on the worm wheel. A double-threaded worm has the lead equal to twice the pitch and will then advance the worm wheel $2/N$ revolutions per turn of the worm. Thus, worm gears follow the general rule of angular velocity ratio inversely proportional to the ratio of the numbers of teeth. Worms are right- and left-hand in the same sense as helical gears. Changing hand of the worm reverses the relative rotation of the worm wheel.

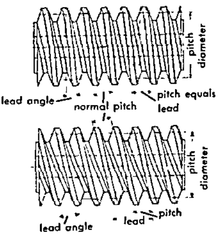


Fig. 9. Characteristics of single and double threaded worms. (*From C. W. Ham, E. J. Crane, and W. L. Rogers, Mechanics of Machinery, 4th ed., McGraw-Hill, 1948*)

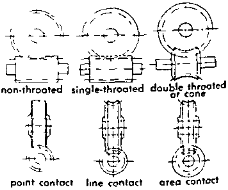


Fig. 10. Nature of contact for worm gears. (*Michigan Tool Co.*)

Improved load-carrying capacity and wear characteristics are made possible by increasing the contact between the worm and wheel (Fig. 10). Line contact is obtained by making the worm wheel surface concave to conform to the tooth profile of the worm. Still greater contact is obtained by using a concave worm as well. Known as a cone-drive or Hindley worm, this design permits greater contact surface and allows more teeth to be in contact at one time. *See* GEAR TRAIN; MECHANISMS; PLANETARY GEAR TRAIN. [R.C.F.]

Bibliography: American Gear Manufacturers' Association, *Allowable Backlash and Composite Error in Gears*, AGMA Standard 236.04; American Standards Association, *American Standard Gear Nomenclature*, ASA B6.10, 1954; American Standards Association, *American Standard Spur Gear Tooth Form*, ASA B6.1, 1932; American Society of Mechanical Engineers, *System for Straight Bevel Gears*, B6.13, 1955; E. Buckingham, *Analytical Mechanics of Gears*, 1949; D. W. Dudley, *Practical Gear Design*, 1954; Fellows Gear Shaper Co., *The Involute Curve and Involute Gearing* and *The Internal Gear*.

Gear cutting

The cutting or forming of a uniform series of toothlike projections on a surface of a workpiece, the teeth being designed to mesh with a mating tooth series in order to transmit power or motion.

Gear-cutting methods may be divided into two general categories. The first is gear generating in which the tooth is produced by the conjugate or total cutting action of the tool plus the rotation of the workpiece. The second method is gear forming in which the desired tooth shape is produced by a tool whose cutting profile matches the tooth form. Frequently gear teeth may be rough cut by a generating method and finished to size by a form tool.

Gear-cutting machines are designed to hold rigidly and to position accurately both a gear blank and a cutting tool in relation to each other so that the designed operation of the machine will produce the desired tooth series on the piece.

Hobbing. Gear-hobbing machines revolve the gear blank in a horizontal plane while the rotating hob or cutter moves downward across the face of the gear to generate the teeth. The hob, similar to a work gear, is made with a spiral thread on its periphery. Gashes across the thread provide the cutting edges. The power arbor, which holds the hob, is tilted so that the cutting teeth line up with the teeth on the gear blank. A system of index gearing maintains the relative turning ratio between the blank and the hob, and changing gear combinations permits the desired number of teeth to be cut. Other gearing or additional mechanism is provided so that the hob and gear blank are either advanced or retarded in respect to each to cut a helix angle on the blank. The arbor must be tilted an additional amount so that the hob thread angle coincides with the helix angle of the gear. The hobbing process is used for roughing as well as for finish cutting of gears.

Generating. Gear-generating machines produce straight-bevel gear teeth by reciprocating a cutting tool across the gear face while a generating motion is attained by slowly rotating the work and the tool holder or cradle at a synchronized rate. The gear blank is automatically withdrawn from the tools at the end of each roll so that both the cradle and work spindle roll back to starting position to cut the next tooth.

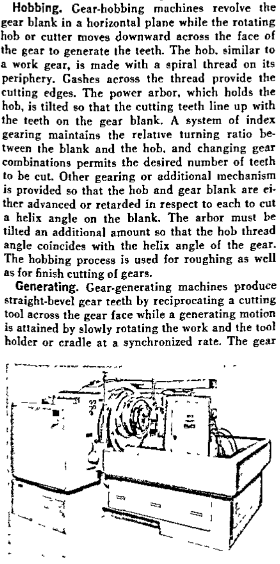


Fig. 1. Hypoid gear generator. (*The Gleason Works*)

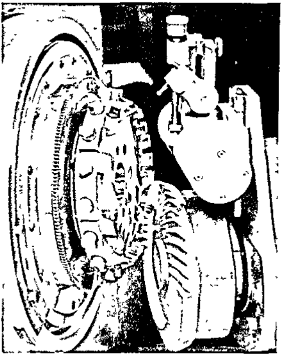


Fig. 2. Gear blank and cutter on hypoid generator. (*The Gleason Works*)

Two rectangular tools are used for cutting. The tools are ground and positioned so that each side of one gear tooth is cut during a generating roll.

Spiral and curved toothed bevel gears and pinions are generated by the cutting action of a face milling cutter synchronized to rotate with a gear blank (Fig. 1). The cutter is a circular tool holder with a series of cutting blades around the edges of its face (Fig. 2). After each generating roll of the tool and blank, the machine indexes to bring the cutter into the next tooth slot.

Planing-type, generating machines, using a single reciprocating tool, are designed to cut large spiral bevel gears.

Shaping. Gear-shaping machines generate gear teeth on a gear blank by means of a pinion-shaped cutting tool. A cutting edge is provided along one side of the teeth. The cutter generates by simultaneously reciprocating across the gear face while it also rolls with the blank. Because the cutter must return to its starting position at the end of each stroke, it is usually relieved on the return stroke by automatically increasing the center distance between work and cutter spindles. Large gears are frequently finish-lapped after shaping. *See* LAPPING.

Gear shaping requires only a small amount of run out at the end of each cutting stroke permitting shaped teeth to be located close to shoulders.

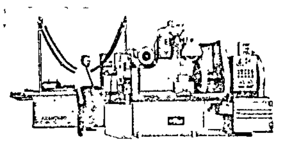


Fig. 3. Automatic gear-grinding machine. (Gear Grinding Machine Co.)

Milling. Gear-milling machines produce gears by use of a rotating cutter whose cutting profile matches the contours of the gear teeth During the cutting process the gear blank remains stationary as the cutter forms both sides of the space between two gear teeth. At the completion of the cut the cutter is returned to its starting position and the blank is indexed or rotated for the next cut. Spur and helical gears, worms, and sometimes bevel gears may be produced by milling. Frequently, standard rather than special machines are used.

Milling does not usually produce gears of great accuracy and often the ...

both by formed grinding wheels and by generation. Although some fine-pitch gear teeth are finish-ground from solid stock, gear grinding is primarily a finishing operation. Grinding is the only way to finish gears which have been brought to a high degree of hardness. *See* GRINDING.

A form-grinding disk wheel grinds both sides of the space between two gear teeth as it moves across the face of the gear (Fig. 3). An involute tooth form is dressed into both sides of the grinding wheel.

Generating grinding is done with a flat-sided disk wheel. As the disk rotates it both reciprocates across the gear face and moves sideways as the gear rotates. Collectively these movements result in an involute tooth form.

Shaving. Gear shaving is used for finishing and for improving profile accuracy. Shaving is accomplished by using both rotary and rack-type cutters whose teeth are serrated with many small notches or cutting edges. The axis of the cutters is set at a slight angle to the axis of the workpiece so that as the two are rolled together the notched cutter teeth shave the gear teeth. *See* BROACHING.

Only gears of machinable hardness are usually shaved, although gears may be shaved before or after they are hardened.

Broaching. Gear-broaching machines are generally used to produce internal gears in addition to racks and gear segments. Small internal gears are frequently made in one pass of the broach. Larger ones require several passes plus indexing of the gear. Either spur or helical teeth may be produced by broaching. *See* MACHINING OPERATIONS. [A.T.]

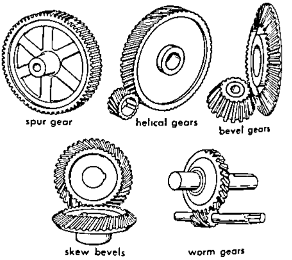
Gear drive

Transmission of motion or torque from one shaft to another by means of direct contact between toothed wheels. The active parts of the gear teeth are usually made in the form of involutes and transmit the force smoothly from a tooth on one gear to a tooth on another gear.

If a 20-tooth pinion (the smaller of a pair of gears) meshes with and drives a 40-tooth gear, the pinion must rotate twice for each revolution of the gear. In general, the ratio of the angular velocity ω of gears A and B is

$$\frac{\omega_A}{\omega_B} = \frac{N_B}{N_A}$$

where N_B and N_A are the numbers of teeth on the respective gears. If friction is negligible, and it



Typical gear drives. (W. H. Crouse, *Automotive Mechanics, 4th ed., McGraw-Hill, 1960*)

usually is in gear drives transmitting power between parallel shafts, input power to A equals output power from B. Thus, expressing power as the product of torque and angular velocity,

$$T_A\omega_A = T_B\omega_B$$

or

$$T_B = T_A \frac{\omega_A}{\omega_B}$$

where T_A and T_B are the torques on shafts A and B respectively. *See* SIMPLE MACHINE. [R.M.PH.]

Gear loading

The power transmitted or the contact force per unit length of a gear. If the gear speed, size, and tooth contour are fixed, the power-handling capacity may be increased by increasing the axial length of the gear. However, where space is restricted (as in aircraft and many ground vehicles) the gear is usually loaded to its safe limit.

To obtain satisfactory tooth-surface durability from highly loaded gears, experience indicates that several items of the gear set must be properly

designed and manufactured, namely, proper modification of the tooth profile to suit the operating conditions; index of teeth and parallelism of teeth, which must be held within close limits; gearing, which must be mounted so the teeth will not deflect out of line; and gear tooth surfaces, which must be of sufficient hardness and proper finish and which should have good lubrication, particularly on start of initial operation.

Gear design is a compromise between tooth strength and surface durability. Larger teeth give more tooth strength but less surface durability; smaller teeth result in less tooth strength but more surface durability.

Surface distress on the teeth becomes particularly important at loadings in the order of thousands of pounds per inch of face.

In designing the gear tooth surfaces, one mesh of teeth is assumed to carry the load through the action, and the action is balanced so that the product of Hertz maximum compressive pressure in lb/in.^{2} and sliding velocity in ft/sec is the same at each end of the assumed single-mesh action. This product (known as a PV value) may be as high as 3,000,000 or higher in gears with high pitch line speeds. The design may produce a smaller contact ratio than conventional, but this result indicates that contact ratio alone is an unreliable measure of gear capacity.

A highly loaded tooth of adequate rigidity deflects about a point in the middle of the rim, bending as a rigid body under load rather than as a nonuniform beam only. Relief of the tooth profile provides clearance for deflection of the preceding mesh, and ramps at the tooth tips assure that first contact does not extend to the tips

This design refinement necessitates corresponding care in production, as in minimizing distortion during carburizing. Teeth can be held parallel within 0.0003 in. in the width of the tooth, and the index can be maintained within 0.0002 in. between adjacent teeth of a gear.

To assure removal of all possible error in grinding and also to achieve a more satisfactory working surface on the gear tooth face, grinding wheels are dressed to produce a finish which is coarser than usual. Slight surface roughness (Profilimeter readings between 15 and 37) results in less surface distress, or scuffing, of highly loaded steel gears than where smoother finishes are used, although a steel gear running with a bronze one should be smoother.

Quiet operation is an indication of efficient operation free from abrupt tooth engagement and severe scuffing of one tooth face by another. Helical gears provide smooth transition of effort from tooth to tooth. The gear is designed with a modified involute to provide soft contact at the start of engagement and end of engagement to reduce the pressure on the teeth when the sliding of the tooth surfaces is highest. For quietest operation they should be cut and shaved accurately for lead, profile, and index. A helical contact ratio of 2 or more is desirable; effort expended to increase [illegible] contact ratio reduces deflection and values of pressure times sliding velocity more than does the same effort expended to increase involute contact ratio. Helix angles of over 45° on constant mesh gears have more expensive mountings but improve quietness. Rigid and accurate mounting is essential to achieve best operating condition.

With other gears subject to deflection, such as hypoid rear-axle gearing, conventional practice is to crown the tooth and modify the tooth profile to keep contact away from the tooth ends and tips. The gear set is usually designed for stiffness rather than for strength, and the shafts and mountings should also be designed on this criterion so that little crowning is needed. [F.R.M.]

Bibliography: J. O. Almen and J. C. Straub, Transmission gears, *Automotive Ind.*, vol. 75, 1937; American Gear Manufacturing Association, *PVT Values of Gear Teeth*, AGMA Progr. Rept. 101.02, 1951; F. R. McFarland, Experiences, highly-loaded gearing, *Proc. Ind. Math. Soc.*, 1959; S. P. Timoshenko, *Theory of Elasticity*, 2d ed., 1951.

Gear train

A combination of two or more gears used to transmit motion between two rotating shafts. In theory only two gears are necessary to give any single speed change desired. This is often not practical, however, because the size of the larger gear would be excessive. For such cases a train consisting of several gears and requiring considerably less space is used to produce the desired velocity ratio. Two general types of gear trains are used. An ordinary train is one in which all gears rotate on fixed axes relative to a single frame. An epicyclic train is an inversion of an ordinary train having at least one gear whose axis moves relative to the frame. Belt, rope, and chain drives are frequently used in conjunction with gear trains. *See* AUTOMOTIVE TRANSMISSION; BELT DRIVE; CHAIN DRIVE; PLANETARY GEAR TRAIN; ROPE DRIVE.

In a gear train, a driver is a gear which, by tooth contact, imparts motion to a driven gear. Gear A (Fig. 1) is a driver, imparting motion to the driven gear B, gear C is a driver imparting motion to gear D, and so on. An idler (gear B, Fig. 3) is both a driving and a driven gear and is used to change direction of rotation. When two shafts are connected by two external gears they rotate in opposite directions. By placing an idler between them, their sense of rotation will be made the same and their velocity ratio will be unchanged. An idler may be used to reduce the diameters of the gears that connect two shafts at a fixed center distance. The size of an idler has no effect on the velocity ratio of the two gears it connects.

The rotation of any gear in an ordinary train is most easily determined by inspection, keeping in mind that mating external gears have opposite directions of rotation and an internal gear has same direction of rotation as its mating gear.

[illegible]lar velocities of two gears in mesh [illegible]

versely proportional to the numbers of teeth on the gears. These two concepts form the basis of analysis of ordinary gear trains.

Ordinary train. In an ordinary gear train (Fig. 1), power is transmitted from the first driving shaft (shaft 1) to the last driven shaft (shaft 4). Although this example is not a compact arrangement, it does illustrate the nature of a gear train. This train shows two examples of compounding—that is, more than one gear fixed to a given shaft. Gears B and C are compounded on and fixed to shaft 2. In like manner, gears D and E are compounded on shaft 3. A simple gear train, by contrast, has each gear carried on a separate shaft in the manner of gears A and F. A 1000-hp rolling mill drive and pinion stand (Fig. 2) illustrates an ordinary train. The first speed reduction is with the opposed single helical gears and the second reduction is through the herringbone gears.

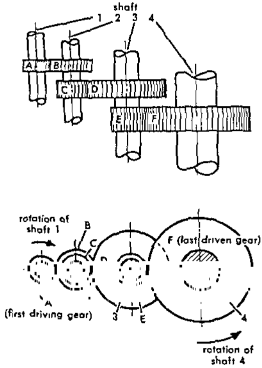


Fig. 1. Ordinary gear train.

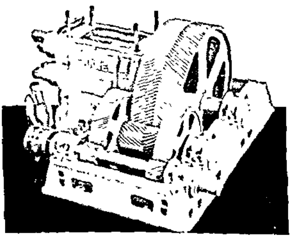


Fig. 2. Helical gear speed reducer. (Farrel-Birmingham Co.)

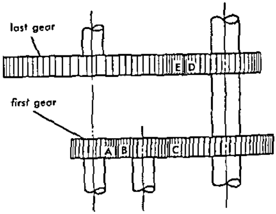


Fig 3. Reverted gear train.

Reverted train. In a reverted gear train (Fig. 3) the first and last gears, while on different shafts, are on the same axis. Automotive transmissions and the back gears of a lathe use the reverted train principle.

Train value. The train value E is, by definition,

$$E = \frac{\text{angular velocity of last driven gear}}{\text{angular velocity of first driving gear}}$$

when the angular velocities are measured with respect to the frame supporting the gears. The velocity ratio of the train is the reciprocal of the train value. Referring to the train of Fig. 1, the train value is

$$E = \frac{\text{rpm shaft 4}}{\text{rpm shaft 1}} = \frac{N_4}{N_1}$$

where N is revolutions per unit time. This may be expanded as an identity for the entire train to

$$E = \frac{N_4}{N_1} = \frac{N_2}{N_1} \times \frac{N_3}{N_2} \times \frac{N_4}{N_3}$$

Each of the ratios on the right side of this equation is the velocity ratio for a mating pair of gears. Letting the number of teeth on any gear be represented by the letter designating the gear and replacing velocity ratios by the inverse tooth ratios gives

$$E = \frac{N_4}{N_1} = \frac{\mathrm{A}}{\mathrm{B}} \times \frac{\mathrm{C}}{\mathrm{D}} \times \frac{\mathrm{E}}{\mathrm{F}}$$

In this equation A, C, and E are the driving gears while B, D, and F are the driven gears. Thus a general expression for the train value is

$$E = \frac{\text{continued product of driving teeth}}{\text{continued product of driven teeth}}$$

Because any two gears in mesh must have the same diametral pitch, expressions for the train value or velocity ratio may also be written in terms of pitch diameters.

Epicyclic train. An epicyclic train derives its name from the path described by any point on one of the gears as it rolls on another. The term epicyclic refers only to the motion and is not related to the gear tooth form, which may be involute or of any other form. Epicyclic trains may be used to obtain considerably greater speed changes than would be possible with an ordinary train of the same size.

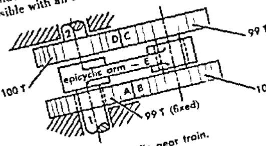


Fig. 4. Reverted epicyclic gear train.

For example, in a compound epicyclic spur gear train (Fig. 4), gear A is fixed and the epicyclic arm E, fixed to shaft 1, carries gears B and C. Analysis to determine the velocity ratio of such a train can be based on a division into two parts of the net motion of the train. In the first part, the train is locked with no relative motion between any of its components. For this analytic step, fixed gear A is released from the frame and locked to gear B. The locked train is rotated one revolution in an arbitrarily selected direction. Clockwise rotation is assumed positive and counterclockwise negative. In the second part, the assembly is considered as an ordinary train with the epicyclic arm becoming the fixed member. While the fixed gear is returned to its original position by one revolution in the opposite sense, the consequent rotation of the other gears is noted. For any one gear, the algebraic sum of the revolutions during the two parts is the net motion for one revolution of the epicyclic arm. To illustrate, consider the following tabulation for the train of Fig. 4 with the number of teeth as marked for each gear. The table shows that, for each clockwise revolution of shaft 1 and arm E, shaft 2 and gear D will turn $^{199}/_{10,000}$ of a revolution in a clockwise direction.

	Gear A	Gear B	Gear C	Gear D	Arm E
Locked train	+1	+1	+1	+1	+1
Locked arm	−1	$+\frac{99}{100}$	$+\frac{99}{100}$	$-\frac{99}{100} \times \frac{99}{100}$	0
Net motion	0	$+\frac{199}{100}$	$+\frac{199}{100}$	$+\frac{199}{10,000}$	+1

Some epicyclic trains have no fixed gear. The analysis of such trains can be made using a modification of the tabular method or by the equation

$$E = \frac{\text{turns of last gear} - \text{turns of arm}}{\text{turns of first gear} - \text{turns of arm}}$$

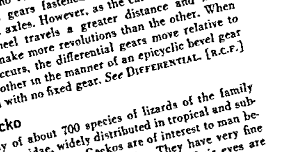
where the train value E is in terms of angular velocities with respect to the epicyclic arm and the first and last gears are of the epicyclic train. The sign convention for rotation must be followed throughout the analysis.

When the tabular method is used for an epicyclic train with no fixed gear, the first step is to give the locked train the same rotation as the known motion of the epicyclic arm. In the second step, with the arm stationary, the gear whose velocity is known is rotated sufficient turns in the proper direction to make its net motion equal the known value.

Probably the most common epicyclic train with no fixed gear is the automobile differential. When the automobile is moving along a straight path, there is no relative motion between the bevel differential gears fastened individually to the right and left axles. However, as the car turns a corner, one wheel travels a greater distance and hence must make more revolutions than the other. When this occurs, the differential gears move relative to each other in the manner of an epicyclic bevel gear train with no fixed gear. See DIFFERENTIAL. [R.C.F.]

Gecko

Any of about 700 species of lizards of the family Gekkonidae, widely distributed in tropical and subtropical climates. Geckos are of interest to man because of several peculiarities. They have very fine scales and their skin appears soft. Their eyes are



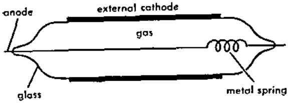
The banded gecko, Coleonyx variegatus; length to 4 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

large with elliptical pupils. Most species have a thick, brittle tail which is easily broken off. The toes of most geckos have enlarged, flattened, lamellated plates covered with hairlike extensions, enabling them to run on smooth surfaces, climb perpendicular walls, or, in some instances, even walk on ceilings.

The geckos occupy a diversity of habitats from the deserts to the tropical jungles, but those species which inhabit houses are best known. They are usually welcome guests because of their insect-eating habit. Their presence is considered a good omen by many people. Most geckos are nocturnal. There are 7 species in the United States, all of them along the southern border; 3 are native species, the other 4 introduced. See SQUAMATA. [J.D.B.]

Geiger-Müller counter

A detector of ionizing radiation. When a fast-moving charged particle traverses a Geiger-Müller counter, an electrical impulse is produced. T



Cylindrical external cathode Geiger-Müller counter. The glass is thin soda glass, and the central wire is 0.003-in. diameter tungsten. The metal spring is used to keep the central wire taut.

impulses can readily be counted by electronic circuits. Geiger-Müller (GM) counters, usually referred to simply as Geiger counters, are widely used to indicate the presence and intensity of nuclear radiations.

Construction and operation. A GM counter consists of a gas between two electrodes (see illustration). One electrode, usually cylindrical and hollow, is the cathode. The other electrode, a fine wire stretched along the axis of the cylinder, is the anode. A potential of about 1000 volts is placed on the wire. When an atom of the gas between the two electrodes is ionized by collision with a charged particle passing through the gas, the electron produced in the collision is drawn toward the central wire The electron then collides with the atoms of the gas. Near the central wire the electric field is very intense, and the electron may acquire enough energy between two collisions to allow it to ionize another atom. A second electron is then set free, and by successive collisions, an avalanche of electrons is produced which is then collected as charge on the central wire. This charge produces an electrical impulse which in typical cases may be 50 volts.

A GM counter is a nonproportional particle detector; that is, the pulse produced is independent of the nature and energy of the particle that liberated the electrons within the gas (*see* PARTICLE DETECTOR). The operation of a GM counter depends critically on the voltage placed on the central wire. If the voltage is too low, the electron avalanche never builds up, and the counter operates only as an ionization chamber (a device which gives the total ionization produced in the gas) or proportional counter, in which the output pulses are much smaller (*see* IONIZATION CHAMBER). If the voltage is too high, additional electrons produced by the positive ions start additional discharges, and the counter discharges continuously. Starting at the threshold for GM counting and continuing up to the region of multiple pulsing is the plateau region of the GM counter. The threshold is rather sharp, and occurs at an anode potential of 800–1000 volts in typical counters. The plateau of a good GM counter extends 100–500 volts above threshold.

The pulses coming from a GM counter are amplified electronically and are then counted by an electromechanical register or as clicks in a loudspeaker. If the counting rate is so high that the mechanical register cannot follow the pulses fast enough, the pulses are fed into a scaling circuit, which divides the number of input pulses by a known scaling factor before they are fed to the register. The pulses may also be sent into a counting-rate meter, which contains a condenser that charges a certain amount each time a pulse reaches it. The accumulated voltage on the condenser indicates the counting rate. This voltage can be measured if a resistance is placed across the condenser to discharge it slowly.

Coincidence counting. GM counters are often used in coincidence. When a single particle passes through two or more GM counters, the pulses from each counter are practically time-coincident. The pulse from each counter is then sent to a coincidence circuit which indicates pulses coincident in time. Coincidence counting is used to eliminate much background count due to local radioactivity. Arrays of GM counters in coincidence also select particles going in a given direction. This technique is used, for example, to measure the angular distribution of cosmic rays.

Types and dimensions. Construction of GM counters requires careful control of both cleanliness of the electrode surfaces and purity of the gas. An inert gas such as argon is often used, mixed with a small amount of organic vapor to a total pressure of about 0.14 atm. The organic vapor quenches the discharge and prevents multiple pulsing. However, some of the vapor is used during each pulse, and such tubes thus have a limited life, usually 10^8–10^9 counts. GM tubes quenched by halogen gas are also available, and these have much longer life. The central wire of a GM counter is usually made of tungsten, 0.003–0.005 in. in diameter. The outer electrode may be copper, brass, or aluminum. For certain applications, external cathode counters, such as the one shown in the illustration, are used. The cathode is formed by spraying a conducting coating on the thin glass envelope surrounding the central wire. The electrical conductivity of soda glass, while small, is sufficient to permit the current to pass through the glass. Such counters have a long recovery time, but are useful in cosmic-ray research and are much cheaper than metallic counters. They have the additional advantage of not being destroyed if an over-voltage is accidentally applied.

GM counters may be made in various sizes, depending on the use for which they are intended. Counters ½-in. in diameter and 3 in. long, with thin aluminum walls, are used for detection of β-particles and cosmic rays. Counters of diameter 2 or 3 in. have been made, with lengths up to 3 ft. However, if a large area is to be covered by GM counters, it is customary to use several connected electronically in parallel. Point GM counters have a somewhat different geometry than cylindrical counters. A sharp point is used as the anode, and a flat plane or sphere is the cathode.

Applications. GM counters are widely used in industry, medicine, mineral exploration, and scientific research.

Industry. A typical industrial use is in the measurement of thickness of sheet material. A source of β-rays is placed on one side of the material and a GM counter on the other. Some of the β-rays are absorbed in the material, and the counter responds to changes in absorption caused by changes in the thickness. Continuous accurate measurement of the thickness can be made even if the sheet material is moving at high speed. Another industrial application is in the detection of traces of radioactive elements in research on the wear of bearings. GM counters are used in connection with x-ray machines to search for flaws in large castings.

Medicine. Radioactive sources used in medicine are often detected with GM counters, which can locate the position or distribution of the radioactive material after it has been administered In medical applications, it is important to determine the total radiation dosage and to monitor the radiation; GM counters are used for this purpose. They have also been valuable in locating expensive and potentially dangerous sources of radioactivity which have been lost.

Geology. Geologists use GM counters in mineral exploration. Minerals containing uranium are radioactive, and can be detected by GM counters. Simple lightweight portable GM detectors are available for prospectors.

Research. Chemists, physicists, and biologists use GM counters in various types of scientific research. Radioactive tracers used in chemical research are followed through complicated chemical reactions by GM counters. GM counters give information on the quantity of material present and also on the lifetime of the element produced in transmutation of elements resulting in radioactive products. GM counters can be made using a radioactive gas instead of the gas normally used In such counters, the wall thickness is not important because the ionizing particle originates within the counter. Applications of GM counters in high-energy nuclear physics are confined to low-counting-rate experiments, such as in cosmic rays, because of the long dead time of GM counters. GM counters are rarely used near particle accelerators because they become jammed in a large flux of ionizing particles. [W.B.F.]

Bibliography: S. A. Korff, *Electron and Nuclear Counters*, 2d ed., 1955.

Gel

A two-phase colloidal system consisting of a solid and a liquid. Gels behave as elastic solids and retain their characteristic shape, whereas sols (colloidal dispersions) possess the shape of the container. Commonly, gels have a low solid content, for example, 2–5% for ferric oxide, and as little as 0.1% for coagulated blood. Gels include the jellies or transparent elastic gels rich in liquid, and gelatinous precipitates which are believed to consist of minute particles of jelly. Gels or jellies which have dried until apparently solid are often referred to as zerogels, and in certain cases, will swell or redisperse to form a sol when treated with a suitable solvent.

The lyophobic gels (usually inorganic) may be prepared by double decomposition reactions, employing high reagent concentrations. Rapid mixing favors the formation of clear, transparent gels. Other methods include slow coagulation of sols by electrolytes and concentration of sols by slow evaporation of the dispersion medium.

The lyophilic gels (usually organic), such as gelatin, agar agar, and certain soaps, may be prepared by cooling a sol prepared at an elevated temperature or by allowing air-dry gels to swell in a solvent. The setting or gelation of a sol is characterized by (1) time of set, (2) gelation temperature, (3) critical concentration of setting, and (4) rate of viscosity increase.

Setting of a sol to a gel occurs with negligible heat effects; this phenomenon is designated as the isothermal sol-gel transformation. Gels such as clays, which contain platelike particles, may be transformed into sols by shaking; these sols revert to gels on standing. This phenomenon is called thixotropy. Rheopexy refers to the converse process, in which gelation occurs more rapidly when the sol is stirred or vibrated

The classical explanations of gel structure were the so-called honeycomb and brush-heap theories. In the former, the continuous phase was the solid, with the liquid contained in the connected holes or pores. In the latter theory, the liquid was the continuous phase, and the pores or capillaries were the interstices between the solid particles. The brush-heap theory appears to be correct for many common gels. For example, one sample of silica gel, as viewed in the electron microscope, consisted of small primary particles about 100 A in diameter. The continuous pore structure (the interstices between the loosely packed primary particles) had a most frequent pore radius of 30–35 A. These considerations do not preclude the possibility that some types of gels, such as sintered or partially sintered gels, may possess the honeycomb structure. *See* COLLOID; GELATIN; LUBRICANT; PECTIN; PRECIPITATION (CHEMISTRY); RHEOLOGY.

[W.O.M.]

Gelatin

A protein derived from the skin, white connective tissue, and bones of animals. It is composed entirely of amino acids joined in polypeptide linkages to form a linear polymer. It gives typical protein reactions and is readily hydrolyzed to yield its amino acid or peptide components. *See* AMINO ACIDS; PROTEIN.

Manufacturing process. The protein collagen is

which is isoionic at pH 4.7–5.0. Frozen edible-grade porkskins, processed to yield type A gelatin, provide about half of the total United States production. Calfskin and beefskin pieces, known

splits, after treatment with concentrated lime solution for periods up to 90 days, yield type B gelatin and provide about one-third of the United States total. Bones of the water buffalo and beef cattle, after demineralization by leaching with dilute hydrochloric acid, are known as ossein. This raw material is usually lime-treated to yield type B gelatin and provides about one-sixth of the United States total. After pretreatment with acid or lime, the raw materials are washed, to remove mineral matter and other impurities, and are extracted with warm water several times to remove the gelatin. The solution is filtered, concentrated in vacuum evaporators, filtered again, and is then chilled to form a sheet of jelly. This is dried and then ground to the desired particle size. To hasten drying, the jelly is sometimes extruded through a die to form noodlelike material with a much greater specific surface, thus facilitating the removal of moisture

Characteristics. The jelly strength of gelatin, expressed in grams, is measured at 6.66% concentration, with the Bloom gelometer, which is a type of penetrometer. Highest grades test 250–325 Bloom, and the weakest grades 50–100 Bloom. Average molecular weight probably varies from a low of 20,000 to a value of 120,000 for the highest grades.

The protective colloidal efficiency, as measured by the Zsigmondy gold number, is 0.009 for type B gelatin and 0.02 for type A gelatin, thus indicating that type B is slightly more effective than type A. The gold numbers are constant over the entire Bloom range for each type. These low gold numbers indicate the very high protective colloidal value of gelatin.

Amino acid composition. The approximate composition of amino acids is shown in the table. Although gelatin is not a complete protein because it lacks the essential amino acid tryptophan, it does supply several amino acids which are lacking in the protein of wheat, barley, and oats. It thus exhibits a protein-sparing action in diets where the cereal grains are a principal component. Pure gelatin is free from antigens and hence does not cause allergic reactions. Where it is desirable to increase the dietary protein, for growing children, for obese persons, and for geriatric feeding, gelatin is an excellent dietary supplement.

Uses of gelatin. The use of gelatin as a food more than tripled in the United States from 1933 to 1958. The principal uses are (1) gelatin desserts; (2) meat products, such as canned hams, meat loaves, luncheon meats, headcheese, scrapple, and souse; (3) candy such as marshmallows, circus peanuts, wafers, fondant; (4) ice cream, sherbets and water ices; (5) canned soups such as jellied consommé, madrilene; and (6) in such bakery items as icings, frostings, cake fillings, and chiffon pie fillings. United States annual usage of gelatin in foods is estimated at 45,000,000 lb (1958).

The use of gelatin in photography is second in volume to food usage and is estimated at 11,000,000 lb annual United States consumption (1958). Because it is a most effective protective colloid, and is easily tanned to render it insoluble, gelatin is the basic ingredient in the manufacture of silver-sensitized photographic film and photographic paper. No other substance has been found that will successfully replace it in this application. In addition to controlling the grain of the silver halide, gelatin is involved in controlling the photographic characteristics, due to trace impurities in the various types of gelatin which act as restrainers, sensitizers, and desensitizers. Gelatin is also used in lithography and in the graphic arts.

Composition of amino acids

Amino acid	Amino acid content, %: Type B gelatin from ossein, splits or calfskin	Type A gelatin from porkskin
Nitrogen	17.4	18.0
Alanine	8.7	9.2
Glycine	26.9	30.5
Valine*	2.6	2.7
Leucine*	3.1	3.2
Isoleucine*	1.9	1.5
Proline	14.0	16.3
Phenylalanine*	1.9	2.1
Tyrosine	0.14	0.69
Tryptophan*		...
Serine	2.9	2.9
Threonine*	2.2	2.2
Cystine	0.05	0.09
Methionine*	0.85	0.80
Arginine*	6.4	8.8
Histidine*	0.63	0.67
Lysine*	5.2	5.1
Aspartic acid	6.9	6.3
Glutamic acid	12.1	11.7
Hydroxyproline	14.4	13.1

* Those listed as "essential" amino acids in human nutrition

Pharmaceutical gelatin (USP) accounts for an annual usage of 5,000,000 lb annually in the United [illegible]

emulsions. A special pyrogen-free grade of gelatin is used as a substitute for blood plasma and for depot medication.

Technical grades of gelatin are used in the chemical industry where a protective colloid is needed to control particle size, prevent phase separation, or to aid in clarification. United States annual usage of technical gelatin is estimated at 5,000,000–10,000,000 lb. *See* FOOD ENGINEERING.

[J.A.D.]

Bibliography: F. L. DeBeukelaer, J. R. Powell, and E. F. Bahlman, Standard methods for determining viscosity and jelly strength of glue, *Ind. Eng. Chem.*, 16:310–315, 1924; B. Idson and E. Braswell, *Gelatin, Advances in Food Research*, 7:235–238, 1957; R. E. Neuman, Amino acid composition of gelatins, collagens and elastins from different sources, *Arch. Biochem.*, 24:289–298, 1949; R. S. Zsigmondy, Die Hochrothe Goldlösung als Regens auf Colloide, *Z. anal. Chem.*, 40:697–719, 1901.

Gem

A mineral or other material that has sufficient beauty for use as personal adornment and has the durability to make this feasible. With the exception of a few materials of organic origin, such as pearl, amber, coral and jet, and inorganic substances of variable composition, such as natural glass, gems are lovely varieties of minerals.

A mineral is defined as an inorganic substance with a characteristic chemical composition and usually a characteristic crystal structure. Each distinct mineral is called a species by the gemologist. Two stones that have the same essential composition and crystal structure but that differ in color are considered varieties of the same species. Thus ruby and sapphire are distinct varieties of the mineral species corundum, and emerald and aquamarine are varieties of beryl. Two or more minerals that have the same structure and are related chemically are called groups. The garnet minerals form a group. *See* MINERAL; MINERALOGY

Most gem stones are crystalline (that is, they have a definite atomic structure) and have characteristic properties, most of which are related directly to either beauty or durability.

Durability. Each mineral has a characteristic hardness (resistance to being scratched) and toughness (resistance to cleavage and fracture).

Cleavage. Cleavage is the term applied both to the tendency toward, and the accomplished fact of, a separation within a gem stone parallel to a crystal face or a possible crystal face.

Fracture. Fracture is a break or a tendency to break in a direction that bears no relationship to the atomic structure of the mineral.

Hardness. The hardness of a gem material determines its resistance to being scratched and thus, in conjunction with resistance to cleavage and fracture, its practicability for use as a jewelry stone.

The hardness of minerals is measured with reference to an empirical scale known as Mohs scale, which consists of 10 minerals, ascending from talc (hardness 1) to diamond (hardness 10). The scale was formulated by determining which of the various well-known minerals scratched others. Gypsum (2) scratches talc, and calcite (3) scratches gypsum, and so on up the scale. However, the hardness of diamond is much greater in relation to corundum (9) than is corundum in relation to topaz (8). With few exceptions, the most important gem stones are those at the top of the hardness ladder; for example, diamond is 10, ruby and sapphire are 9, chrysoberyl is 8½, and topaz, beryl (emerald and aquamarine), and spinel are 8.

Beauty. Optical properties are particularly important to the beauty of the various gem materials. The important optical properties include color, dispersion (or "fire"), refractive index, and pleochroism. Gem stones usually are cherished for their color, brilliancy, fire, or one of the several optical phenomena, such as the play of color of a fine opal or the star effect in a sapphire.

Brilliancy. Brilliancy depends on the refractive index, transparency, polish, and proportions of a cut stone. Refractive index is a measure of a gem's ability to reduce the velocity of light that passes through it. It is defined either as the ratio of the velocity of light in air to the velocity of light within a gem stone, or as the ratio of the sine of the angle of incidence to the sine of the angle of refraction. Although diamond does not have the highest refractive index known, it has one of the highest to be found among transparent materials. This, combined with its unsurpassed hardness (which permits it to be given a magnificent polish) and great transparency, gives diamond the highest brilliancy potential of any mineral or manmade material. Although synthetic rutile (titania, or titanium dioxide) has a higher refractive index, its lower transparency, relative softness (which makes a superior polish impossible), and other factors combine to reduce its brilliancy well below that of diamond.

In order for a gem stone to display its latent brilliancy to best advantage, it must be cut to the proper angles for its refractive index. For example, the correct angle of the pavilion (bottom portion) of a brilliant-cut diamond is 41°, whereas the stones of lower refractive index must be cut to slightly greater angles (measured from the plane of the girdle).

Refractive indices. The refractive indices of the important gems vary from a low of approximately 1.45 for opal to a high of 2.6–2.9 for synthetic rutile The value for diamond is 2.42. Refractive index is a rough measure of brilliancy; in other words, the brilliancy of two colorless stones of equal transparency and cutting quality is approximately in proportion to their refractive indices. The refractive indices of other important gems are as follows: ruby and sapphire, 1.76, spinel, 1.72; chrysoberyl, 1 74; topaz and tourmaline, 1.62; beryl, 1.58.

Dispersion. Dispersion is the breaking up of white light into its component colors. It is measured by the difference in the refractive indices of certain of the red and blue wavelengths of light as they pass through a gem. In order to give comparable figures for dispersion, two specified wavelengths in the red and blue ends of the sun's spectrum are used. For example, the refractive indices for these two wavelengths in spinel are 1.710 and 1.730, respectively; therefore, the amount of dispersion is expressed as 0 020. A comparable figure for synthetic rutile is 0.330.

Selective absorption. Although diamond and colorless zircon are valued for their brilliancy and prismatic fire, gem stones are more often cherished for the loveliness of their colors. The color of most gem stones is caused by selective absorption. Selective absorption refers to a gem's ability to absorb or transmit certain wavelengths of light more readily than others. The wavelengths transmitted with least absorption are those that give a gem its color. Absorption is usually caused by very small per-

centages of metallic oxides that are present as impurities.

Pleochroism. Some gem stones exhibit pleochroism, the property of some doubly refractive materials of selectively absorbing light unequally in the directions of transmission, resulting in color differences. If two different colors are exhibited, the result is called dichroism (displayed by ruby, sapphire, and emerald); if three, the result is called trichroism (displayed by alexandrite).

Asterism. Some of the most highly prized gem stones are those that depend on unusual optical effects for their beauty. Perhaps the most important of these are the star stones, which display the phenomenon called asterism. In a star, the reflection of light from lustrous inclusions is reduced to sharp lines of light by a domed cabochon style of cutting. The usual star effect is the six-rayed star seen in star sapphires and rubies. Four-rayed stars occur in garnets and some spinels. Other six-rayed stars are occasionally seen in such stones as quartz and beryl. In corundum, the most important species in which asterism is a significant feature, the phenomenon is caused by reflection from three sets of needlelike inclusions arranged in planes perpendicular to the sides of the hexagonal (six-sided) crystal, with each of the three sets of inclusions in planes parallel to a pair of faces of the crystal. *See* CORUNDUM.

Chatoyancy. Another important optical phenomenon is chatoyancy, or a cat's-eye effect. The most important gem in which it is a prominent feature is the variety of the mineral chrysoberyl known as precious cat's-eye. This effect, when seen in chrysoberyl, is usually much more silkily lustrous than in any other gem stone. As a result, the term precious cat's-eye has come to be applied to the finer specimens of the chrysoberyl variety. Other gem stones in which cat's-eyes are sometimes encountered include tourmaline, quartz, beryl, scapolite, diopside and some of the other rarer gem minerals.

Play of color. The gem mineral opal is one of the most attractive of the phenomenal gems. The colorful display called play of color results from light interference in thin layers of opal that have refractive indices differing from those in adjoining portions.

Adularescence. Another optical phenomenon that produces an interesting gem variety is adularescence, the billowy light effect seen in adularia or in moonstone varieties of orthoclase feldspar. In its most attractive form, it has a slightly bluish cast. Moonstone and opal are among the softer of the fairly common gem minerals, and have hardnesses of approximately 6 on Mohs scale.

Precious and semiprecious stones. Gem stones are commonly designated as precious or semiprecious. This is a somewhat meaningless practice, however, and often misleading, since many of the so-called precious gem varieties are inexpensive and many of the more attractive varieties of the semiprecious stones are exceedingly expensive and valuable. For example, a piece of fine-quality jadeite may be valued at approximately $1000 per carat, whereas a low-quality star ruby may be worth $1 per carat or less. Fine black opals, chrysoberyl cat's-eyes, and alexandrites are often much more expensive than many of the colors of sapphire. *See* CARAT.

Occurrence. Gem stones occur under a variety of conditions in nature. A number of the highly priced ones occur as primary constituents in igneous rocks (those that result from the cooling of molten rock masses) or in alluvial deposits derived from such deposits. The only known occurrences of diamonds in primary deposits are in the ultrabasic rock known as kimberlite, a type of peridotite. Although most of these are volcanic necks or conduits (so-called pipe deposits), dikes and sills of kimberlite also contain diamonds. Diamond-bearing pipes are known in South Africa, Tanganyika, the Belgian Congo, Sierra Leone, India, and Arkansas. They have been reported in Russia as well. Extensive alluvial deposits are mined also in Ghana, Angola, French West Africa, French Equatorial Africa, South-West Africa, Venezuela, and Brazil.

Perhaps the most important types of primary deposits in which gem stones other than diamond are found are pegmatite dikes and contact metamorphic deposits. The intrusion of molten rock into impure limestone often brings about the crystallization of calcite to form a coarse-grained marble with a broad assemblage of other minerals. Frequently, some of the minerals produced in the contact zone include gem-quality crystals such as corundum and spinel. The Mogok area of Burma is the foremost supplier of both the finest rubies and the finest sapphires. The gems in this area are mined from both original contact metamorphic deposits and from gravels derived from those deposits. The deposits of Burma produce an amazing variety of gem-quality materials. In addition to both of the major corundum varieties, other colors of sapphire are also found in this country, as well as peridot, topaz, quartz, garnet, and a wide variety of rarely encountered gem minerals such as scapolite and enstatite.

Pegmatite dikes are found in a number of places throughout the world, but only a few of them produce gem-quality crystals. The most important gem pegmatite areas include Brazil, Madagascar, and Southern California. Minerals recovered from pegmatite dikes include most varieties of beryl other than fine emerald, the kunzite variety of spodumene, topaz, and tourmaline. Fine emeralds occur in calcite veins in a shale in the famous mine near Muzo, Colombia, which had been worked for an unknown period even before the Spanish conquerors stumbled upon it in the sixteenth century. Elsewhere most emeralds occur in mica schist, sometimes with other beryllium minerals such as chrysoberyl.

Most of the important colored-stone mining is done in alluvial deposits. For example, the island

of Ceylon, long known as the "island of gems," has produced a variety of gem stones from alluvial deposits for many years.

Identification. Since most gems are varieties of mineral species, and since mineral species have fairly constant chemical compositions and characteristic crystal structures, their physical and optical properties vary within rather narrow limits. Thus, identification is a matter of measuring these properties and other characteristics, the most important of which are refractive index, optic character, the nature of inclusions, and specific gravity. Because less than 10% of all known mineral species have varieties that are useful as gems, comparatively few tests are needed to separate them; only the first three tests are required in the majority of identifications.

Synthetic gem materials. The properties of synthetic gem materials are identical to those of their natural counterparts; therefore, it is usually necessary to study suspected stones under magnification to determine whether their growth characteristics are natural or the result of synthetic production. For example, the accumulation lines in the form of color banding in natural sapphires are arranged in a hexagonal pattern, whereas those in synthetic materials are curved. Spherical gas bubbles are characteristic of synthetics, but the inclusions in natural sapphires are angular. These same conditions also hold true for ruby and synthetic ruby. To make this determination, binocular magnification with dark-field illumination is the most effective (Fig. 1).

Fig. 1. The Gemscope, a binocular magnifier with dark-field illuminator for gem testing. (*Gemological Institute of America*)

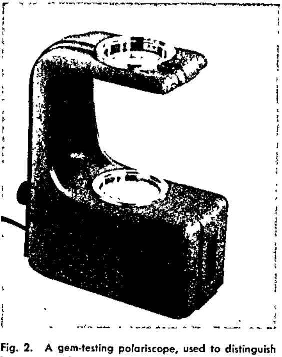


Fig. 2. A gem-testing polariscope, used to distinguish between single and double refraction. (*Gemological Institute of America*)

Refractive index. The refractive index is measured on a refractometer, an instrument that measures the critical angle between an optically dense hemisphere and the gem stone being tested. The instrument is calibrated directly into refractive indices, and as a result, readings from the scale are not in angles but in refractive indices. Doubly refractive minerals show two readings, if the birefringence is sufficiently pronounced (birefringence is the difference between the two indices of a doubly refractive material). Thus, a characteristic birefringence may be read as well as a characteristic refractive index. For example, tourmaline shows two readings, one at 1.624 and another at 1.644, which alone are sufficient to identify this gem mineral. If monochromatic sodium light is used, instead of ordinary incandescent illumination, it is possible to read indices accurately to the third decimal place. When two readings are seen, the necessity for the use of the polariscope, the instrument used to distinguish between single and double refraction, is obviated. The polariscope (Fig. 2) utilizes two polaroid plates, set so that their vibration directions are at 90° to one another, and the gem stone is examined between the two plates. A singly refractive gem stone rotated between the plates remains uniformly dark, whereas a doubly refractive stone becomes alternately light and dark. Diamonds, garnets, and glass remain dark in the polariscope, whereas emerald, ruby, tourmaline, topaz, and zircon become alternately light and dark.

Specific gravity. The density of a gem material compared with an equal volume of water is measured either by heavy liquids or by weighing the stone in air and in water on a sensitive balance adapted for this purpose. Gem materials vary in specific gravity from amber, at about 1.05, to hematite, at 5.20.

Fluorescence characteristics in ultraviolet light, spectroscopy, x-ray diffraction, chemical tests, and hardness and other tests may be helpful in identification.

Gem materials. More than 100 natural materials have been fashioned at one time or another for ornamental purposes. Of these, however, only a relatively small number are likely to be encountered in jewelry articles. The following paragraphs describe the important gem materials (see table for hardness, specific gravity, refractive indices).

Amber. Amber is a yellow to brown fossil resin that is used mostly for beads, pipe stems, cigarette holders, and other items. See AMBER.

Beryl. Beryl is a transparent beryllium-aluminum silicate. When green, it is known as emerald; when light blue, as aquamarine; when pink to red, as morganite. It also occurs in yellow, light green, brown, orange, or colorless forms See BERYL; EMERALD.

Chrysoberyl. Chrysoberyl is a transparent to translucent beryllium aluminate. Its varieties, known as cat's-eye and alexandrite, are among the important gem stones. Cat's-eye exhibits a band of light across the dome of a cabochon, which is caused by reflections from parallel silklike inclusions. The color is greenish yellow to yellowish green. Alexandrite changes color from green in daylight to garnet red in candlelight. *See* CHRYSOBERYL.

Hardness, specific gravity, and refractive indices of gem materials

Gem material	Hardness	Specific gravity	Refractive index
Amber	2–2½	1 05	1 54
Beryl	7½–8	2 67–2 85	1 57–1 58
Chrysoberyl	8½	3 73	1.746–1 755
Corundum	9	4 0	1 76–1 77
Diamond	10	3 52	2 42
Feldspar	6–6½	2 55–2 75	1 5–1 57
Garnet			
Almandite	7½	4 05	1 79
Pyrope	7–7½	3 78	1 745
Rhodolite	7–7½	3 84	1.76
Andradite	6½–7	3 84	1 875
Grossularite	7	3 61	1.735
Spessartite	7–7½	4 15	1.80
Hematite	5½–6½	5 20	
Jade			
Jadeite	6½–7	3 34	1.66–1.68
Nephrite	6–6½	2 95	1.61–1 63
Lapis lazuli	5–6	2 4–3 05	1 50
Malachite	3½–4	3 340–3 950	1 66–1.91
Opal	5–6½	2 15	1 45
Pearl	4	2 7	
Peridot	6½–7	3 34	1.654–1.690
Quartz			
Crystalline	7	2 65	1 54–1 55
Chalcedonic	6½–7	2 60	1.535–1.539
Spinel	8	3 60	1 72
Spodumene	6–7	3 18	1 66–1 676
Topaz	8	3 53	1 61–1.62
Tourmaline	7–7½	3 06	1 624–1 644
Turquois	5–6	2 76	1 61–1 65
Zircon			
Blue and colorless	7½	4 7	1 92–1.98
Green	6	4 0	1 81

Coral. The gem material coral is an assemblage of colonies of the tiny marine animal, coral. It is usually an orange to flesh color and semitranslucent to opaque. *See* CORAL.

Corundum. Corundum is transparent to translucent aluminum oxide. Ruby is the orange-red to violet-red variety; all other colors are called sapphire. The principal colors of sapphire other than blue are yellow, colorless, orange, pink, purple (amethestine), and green. Both ruby and sapphire occur with a beautiful six-rayed star effect. *See* CORUNDUM; RUBY; SAPPHIRE.

Diamond Diamond, the transparent form of carbon, is the most important gem stone. In addition to the popular colorless form, it occurs in yellow, brown, pink, blue, and green. Bombardment by subatomic particles (followed by heat treatment for some colors) yields green, yellow, orange, and brown; blue is also possible. *See* DIAMOND.

Feldspar. The gem stone of importance in the feldspar group is moonstone. It is a semitransparent form of potassium feldspar that exhibits a floating light effect. Less important varieties include the amazonite variety of microcline and the iridescent, translucent gray labradorite. Since several minerals are included in the feldspar group, the properties are variable. *See* FELDSPAR; LABRADORITE; MICROCLINE; ORTHOCLASE.

Garnet. Within the garnet group are several distinct minerals that share the same crystal structure but that differ in properties and chemical composition (all are silicates of aluminum and two other metals). Two of the garnet species, almandite and pyrope, have the appearance commonly ascribed to garnet. A mixture of the almandite and [illegible] more transparent gem stone, with a distinctive violet-red color. Other garnet species include andradite, of which the lovely green demantoid is a variety; grossularite, of which the orange-brown hessonite is a variety; and spessartite, which has varieties closely resembling some of the hessonites. *See* GARNET.

Hematite. The principal ore of iron, hematite, sometimes occurs in dense, hard masses that take a very high polish; these are often fashioned into intaglios. Hematite is metallic grayish black in color. *See* CAMEO; INTAGLIO (GEMOLOGY).

Jade. The material popularly known as jade may be a variety of either jadeite or nephrite. They are characterized by exceptional toughness, which makes even delicate carvings [illegible], and by their [illegible]ring and trans[illegible] The loveliest and m[illegible]le of the twc[illegible] a member

of the pyroxene group of minerals; it is a sodium-aluminum silicate. It is semitransparent to semitranslucent and may be intense green, white with green streaks or patches, brown, yellow, orange, violet, or pink. Nephrite is translucent to opaque and is a darker, less intense green than jadeite; it may also be white, gray, or black. *See* JADE.

Lapis lazuli. Lapis lazuli is an opaque, intensely deep-blue gem stone flecked with golden-yellow pyrite. It has been used for many centuries and is undoubtedly the stone to which the term "sapphire" was first applied. Pliny, for example, described sapphire as an intensely deep-blue stone flecked with gold. *See* LAZURITE.

Malachite. The colorful, opaque mineral malachite is often banded in two or more tones of green and may show a radial fibrous structure. It is always green and is often accompanied by the deep violet-blue azurite, another copper mineral. Malachite is used mostly for cameos and intaglios and for inexpensive scarab bracelets. *See* AZURITE; MALACHITE.

Opal. The fabled beauty of opal is due to a form of light interference in which patches of intense colors are seen against either a white or a nearly black background. There is also a transparent orange to red variety called fire opal. Since it is a particularly fragile material, opal must be treated with care. *See* OPAL.

Pearl. Oriental pearls are those found as lustrous concretions in one of the three species of the salt-water mollusk genus *Pinctada.* Concretions in edible oysters are without pearly luster and are valueless. Pearls found in several genera of fresh-water clams are called fresh-water pearls, to distinguish them from Oriental pearls. Pearls usually occur in white, cream, or yellow colors with rose or other overtones, although black and gray are also very desirable. *See* PEARL.

Peridot. Peridot *is a yellowish-green to green* variety of the mineral group olivine; it is a magnesium-iron-aluminum silicate. In gem quality, it is always transparent and olive green. *See* OLIVINE.

Quartz. In its two major types, single crystal and cryptocrystalline (or chalcedonic), quartz, the most common mineral, has more gem varieties than any other mineral species. The important varieties of crystalline quartz are amethyst, citrine, aventurine and tiger's-eye; the most important varieties of cryptocrystalline quartz are carnelian, sard, chrysoprase, bloodstone, agate, and onyx. Amethyst is purple to violet and transparent. Citrine is yellow to brown and also transparent; it is more commonly known as topaz-quartz and, unfortunately, is often sold as a topaz. Aventurine is usually green with lustrous or colored spangles; it is translucent. Tiger's-eye is a translucent, fibrous, broadly chatoyant, yellow-brown stone that may be dyed other colors. Carnelian is red to orange-red, and sard is a darker brownish red to red-brown; both are translucent. Chrysoprase is light yellowish green. Bloodstone is dark green with red spots. Agate is composed of curved bands in a variety of colors. Onyx is similar to agate, except that the bands are straight. *See* AGATE; AMETHYST; ONYX; QUARTZ.

Spinel. This gem mineral is of particular interest because of the strong resemblance of many of its varieties to comparable colors of corundum. In general, red spinel is less intense in color than ruby; similarly, blue spinel is less intensely colored than blue sapphire. On the other hand, some varieties of spinel are lovely in their own right. Flame spinel is an intense orange-red and is a very attractive gem stone; spinel also occurs in green and amethystine colors. All gem varieties of spinel are transparent, with the exception of the rare black star spinel. *See* SPINEL (MINERAL).

Spodumene. Spodumene is a member of the pyroxene group of minerals; in contrast to jadeite, however, it is fragile. The principal variety for gem purposes is kunzite, which is a lovely light red to light purple transparent stone. *See* SPODUMENE.

Topaz. Topaz is best known in its yellow to brown variety, but the red, pink, and blue varieties are also attractive and desirable. *See* TOPAZ.

Tourmaline. Although the color range of tourmaline is as wide as that of any gem material known in nature, its best-known varieties are red (rubellite) and dark green; it may also be colorless, yellow, blue, black, brown, or other colors. *See* RUBELLITE; TOURMALINE.

Turquois. Turquois is an opaque gem stone with a light intense blue color. Its intense color has attracted mankind from the earliest times. *See* TURQUOIS.

Zircon. Zircon is best known as a transparent colorless or blue gem stone. The colorless variety has been used principally as an inexpensive substitute for diamond. It also occurs in green, yellow, brown, red, and flame colors. The properties of zircon vary rather widely. *See* GEM, MANUFACTURED; GEM CUTTING; GEM MOUNTING; JET (GEMOLOGY); ZIRCON. [R.T.L.]

Bibliography: E. H. Kraus, C. B. Slawson, *Gems and Gem Materials,* 5th ed., 1947.

Gem, manufactured

The term manufacture as used here does not include such processes as shaping, faceting, and polishing, but only the processes that affect the material from which the finished gem is produced. These processes fall naturally into three groups: (1) processes that change the mineral in some fundamental characteristic such as color; the result is called a treated gem; (2) processes by which man makes a material that is identical with the naturally occurring mineral; the result is called a synthetic gem; and (3) processes that produce imitation or simulated material with the appearance of the natural gem; the product in this case is known as an *imitation gem.*

Treated gems. Few gems, except the most expensive, such as diamond, ruby, and emerald, are marketed in naturally occurring colors. Most have l treated to improve or change their color, and

the new color is one not found in the natural material. *See* GEM.

The earliest method extensively used was dyeing and staining, and agate was then and still is today the most widely used gem subjected to this treatment. Colored substances are introduced into the minute interstices between the microscopic quartz crystals. Because of the varying porosity in the individual layers, the banded structure is intensified. The best method of coloring a given specimen can be discovered only by the trial and error treatment of a small sample.

In general the process of dyeing an agate is first to dry the specimen thoroughly; second, to soak it in an appropriate liquid until the pores are penetrated; and finally, to fix the dye within the pores. When the process was first used in the Idar-Oberstein district of Germany, honey or sugar was allowed to penetrate the agate, which was then immersed in sulfuric acid which charred the honey or sugar to carbon. The layers of the agate that had absorbed the sugar or honey developed a brown to black color, depending on their porosity. Some layers with no porosity remained white. Subsequently iron and copper salts were used which were fixed by appropriate chemicals. Organic dyes have also been used.

Heat treatment. This process is used to improve the natural color of a gem or to develop new colors. Amethyst and green tourmaline develop a deeper color when heated, and the brown color of smoky quartz may often be changed to the yellow or reddish-yellow citrine. Some yellow or brownish zircons change to blue, green, or red when heated in a reducing atmosphere or become colorless if heated in an oxidizing atmosphere.

Radiation. Radiation induces many color changes in gems. The induced color varies with the original material and with the type and intensity of the radiation used. Fading in sunlight, characteristic of rose quartz, is an example of a change produced by radiation. Long exposure of colorless quartz to ultraviolet rays will develop a smoky-brown color. X-rays, radium, and other forms of nuclear radiation may produce striking changes in color. Diamond develops first a green color, and with more intense nuclear radiation the green passes to black [illegible]

term metamict minerals is used to describe crystals in which structural changes have been introduced by intensive radiation.

Synthetic gems. The U.S. Federal Trade Commission has restricted the term synthetic gem to manufactured materials that have the same chemical, physical, and optical properties as their naturally occurring counterparts. Many gem materials, including diamond, have been synthesized, but in such small crystals or with such poor quality that they are unsatisfactory as gem stones. Attempts to make gem stones have resulted in producing substances hitherto not known, many of which are of great importance industrially. Others have resulted in significant improvements in existing processes.

In 1891, E. G. Acheson, while attempting to synthesize diamond, produced SiC (Carborundum), the second hardest known substance. Knowledge gained in attempting to make ruby and sapphire aided in the development of nongem crystals of corundum (sapphire, Al_2O_3), which with SiC form the basis of the abrasive industry. The value of synthetic ruby used as jewel bearings in watches and electrical instruments greatly exceeds that used as gem stones.

Early attempts to make synthetic rubies were only partially successful. In 1902, A. V. L. Verneuil, a Frenchman, made ruby and sapphire of gem quality by a process that bears his name; it is also called the flame-fusion method. J. B. Hannay (1880) in England, Henri Moissan (1893), W. Crookes (1906), and others made claims for the successful synthesis of diamond, but these were later disproved. On February 15, 1955, the General Electric Company announced the first proven successful synthesis of diamond but not in crystals of gem quality or size. On October 22, 1957, General Electric started the production and sale of an industrial diamond grit for abrasive use, averaging 1/10 millimeter in diameter, in competition with natural diamond grit which is made by crushing low-grade diamond crystals. The abrasive use of diamond grit for the manufacture of bonded diamond grinding wheels greatly exceeds in volume but not in value the amount of diamond used for other industrial uses and for gem stones. The production of gem quality diamond by the General Electric process is highly improbable. *See* DIAMOND.

Verneuil or flame-fusion process. This process uses the high temperatures produced by an oxy-hydrogen flame.

1. Sapphire. Sapphire is produced by the fusion of very finely divided and highly pure Al_2O_3 which is carried by the oxygen gas in an apparatus that resembles an oxy-hydrogen torch with the flame directed downward into an insulating chamber. The powder fuses into droplets which initiate the development of the cylindrical boule when the flame impinges on a fire clay support. The diameter of the boule is controlled by lowering the fire clay support while regulating the gas flow and the amount of Al_2O_3 introduced. The upper surface of the boule remains molten. Various colors are produced by the introduction of appropriate metal oxides. It is essential to exclude traces of sodium which prevent the formation of clear boules. If 20 mole % magnesium oxide, MgO, is added, a cubic synthetic spinel of fine quality develops instead of the hexagonal corundum boule. Synthetic spinels with variable compositions up to as much as 80% MgO may be made.

The average boule is about ½ in. in diameter and 1–2 in. long, and weighs about 125 carats. Crystallographically oriented boules may be produced by starting with a properly oriented seed

crystal. Rods of clear Al_2O_3 as small as 1 mm in diameter and up to 18 in. long are made by the Linde Air Products Company by a semiautomatic process. After being ground to a uniform diameter, several rods are held firmly together in a bundle and sawed perpendicular to their length by a diamond saw. Each saw cut produces a circular blank from each rod, which is processed into a jewel bearing. From mushroomlike boules as large as 4 in. in diameter, circular disks are cut and polished. Because of the resistance of Al_2O_3 to high temperatures, these disks are used as windows in furnaces. They are also highly conductive in the infrared region of the spectrum. Because of this conductivity, the rods may be inserted into furnaces to transmit the infrared radiation into heat-controlling apparatus on the outside of the furnace.

2. Ruby. Ruby is used for jewel bearings in preference to the clear sapphire because colored jewels are easier for the workmen to handle. Ruby is made by adding 5–6% chromium(III) oxide, Cr_2O_3, to the Al_2O_3. Where extreme hardness is not essential the easier worked spinel is preferred. Colored varieties of spinel are easier to make and give clearer crystals than corundum for all colors except red. The hardness of spinel, 8 on Mohs scale, is adequate for all gem purposes. Synthetic spinel boules are elongated cubes with rounded corners.

3. Rutile. For the production of synthetic rutile (titania) the flame-fusion apparatus is modified by the addition of a third tube at the orifice surrounding the other two through which oxygen passes to oxidize the titanium more completely. The boules, as produced, are black and must be heated in a furnace in an oxygen atmosphere at 1400–1500°C to remove the color. Unless metallic coloring oxides have been added, the boules retain a faint tinge of yellow after the oxygen furnace treatment, because of an absorption band in the blue. The unoxidized titanium atoms in undecolorized boules make them semiconductors whose conductivity is proportional to the amount of color remaining. Synthetic titania boules form elongated tetragonal prisms with rounded corners.

Identification. Under the microscope, synthetic crystals may be distinguished from the natural ones. They are purer than the natural and show only round, included gas bubbles while the natural have minute crystal plates of hematite (Fe_2O_3) and crystalline needles of rutile (TiO_2). Natural crystals have irregularly shaped cavities enclosing both liquid and gaseous inclusions. Curved microscopic bands of uneven pigmentation parallel to the boule outline resemble the fine hexagonal twinning patterns seen on some natural crystals. Both may be confused with the polishing striations that often are microscopically visible on facets of gem stones.

Synthetic stars are sharper than the natural stars and often have areas of clear transparent ruby and sapphire. They are made by adding an excess of TiO_2 to the Al_2O_3 powder and then heating the finished boules at a temperature above 1000°C. The TiO_2 recrystallizes as needles oriented by the host crystal, in the plane of the basal pinacoid, in a hexagonal pattern which is controlled by the crystal structure of the Al_2O_3.

Synthetic rutile boules are clear and transparent, larger than natural rutile crystals and much superior in quality. Natural rutile is always brownish even when transparent.

Emerald. Carrol Chatham of California has developed a fusion process for making emeralds, the details of which are a closely guarded secret. Only a small percentage of the crystals are of gem quality and size. They fluoresce with a deep red color under ultraviolet light, while natural emerald does not.

Imitation gems. Since prehistoric times, glass has been the most widely used imitation gem. Since World War II, colored plastics have replaced glass to a great extent and expanded the use of imitations for costume jewelry. Strontium titanate made by the flame-fusion method has, since 1955, been the most successful imitation of diamond. It has an index of refraction nearly the same as that of diamond and a somewhat higher dispersion, which gives it a brilliancy equal to and a fire (dispersion) somewhat greater than diamond. Synthetic rutile has a higher index and a much higher dispersion than diamond. Its strong dispersion and very high double refraction serve to distinguish it from diamond. Strontium titanate, with a hardness of 5–6, scratches easily; while rutile, with a hardness of 6–7, is somewhat more satisfactory as an imitation gem.

Brilliants and rhinestones made of flint glass are often silvered on the back to give them a high brilliancy. Glass with a high dispersion, which is often called paste diamond, has a high lead-arsenic content. The term paste owes its origin to the practice of wet-mixing the ingredients before firing. It is somewhat softer than ordinary glass. The cheaper forms of glass are cast in molds and, under a hand lens, show rounded edges between the intersections of facets. The better grades, known as cut glass, have been cut and polished after first being molded approximately into the desired form. Cut glass has facets that intersect in sharp edges.

Modern plastics have made costume jewelry very popular not only because they are cheap but because they are lighter in weight and are easily molded or worked into various forms. Thermosetting plastics, like bakelite, were first used in jewelry early in the twentieth century. Unlike thermoplastics, which can be resoftened by heating after molding, thermosetting plastics become permanently hard after the initial molding. The various types are marketed under numerous trade names. *See* PLASTICS FABRICATION. [C.B.S.]

Gem cutting

The polishing of rough gem materials into faceted or rounded forms for use in jewelry is called lapidary or gem cutting. The term lapidary

in application to the cutting of colored stones. One who fashions colored stones is called a lapidary, lapidarist, or lapidist.

Commercial gem cutting. Commercial colored-stone and diamond cutting are separate and distinct fields. Mechanically, the problems of colored-stone cutting are made simpler by the fact that all colored stones are soft enough to be shaped readily using silicon carbide or alumina powder as an abrasive. In contrast to diamond, they may be sawed or polished in any direction. In commercial practice, the gem cutter determines by eye the angles to which facets should be cut in relation to the girdle plane or the top facet. In addition, he creates by eye the perfect symmetry one sees when viewing the finished stone from above. In this respect, commercial gem cutting differs markedly from that done by hobbyist gem cutters, for the hobbyist usually controls the angles and the symmetry by the use of various mechanical faceting heads.

The first step in the usual cutting operation is to saw the crystal or piece of massive gem material to obtain the size and general shape desired. If the material is relatively inexpensive, it may be ground to shape rather than reduced by cutting. Next, if the stone is soft enough, it is usually ground into the approximate final shape against a silicon carbide wheel; if it is more than 7 on Mohs scale of hardness, a diamond-charged metal lap, or wheel, is used. (*For a discussion of hardness of gem stones see* GEM.) Facets are ground and polished on other laps. The commercial cutter often employs what is called the jamb-peg method In this method, a square or rectangular board containing rows of partly drilled holes is clamped into a position at right angles to the spinning lap. The stone that is being faceted is mounted with adhesive in the end of a stick, called a dopstick, the opposite end of which is pointed. The cutter holds the pointed end of the stick in a hole he chooses as one representing the proper angle for the facet being ground. The end of the stick on which the stone is mounted is held against the spinning lap. The term dop is used both in diamond and colored-stone cutting.

In contrast to commercial cutting, hobbyist cutting is much more mechanized. Many faceting heads are on the market for hobbyists. On these, the stone is mounted on a dop and the dop fitted into position on the faceting device. The angles can be set simply, and, in addition, the position of the facet can be indexed so that as many similar facets as necessary can be placed symmetrically around the stone.

Facet grinding is usually accomplished on copper, iron, or lead laps, using diamond or silicon carbide powder for ruby and sapphire or their synthetic counterparts, and synthetic corundum powder for the softer materials. Polishing is accomplished on tin, wood, or plastic laps, with the wooden laps usually covered with leather or cloth. The polishing agents used include tin oxide, cerium oxide, alumina, and diamond powder. Both polishing and grinding laps are operated at rates of speed that are more rapid for hard materials than for soft ones.

Few lapidaries appreciate the importance of the direct relationship between the proportions of the

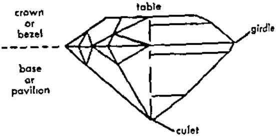


Fig. 1. Arrangement of faceted cuts showing brilliant cut (left side) and step cut (right side).

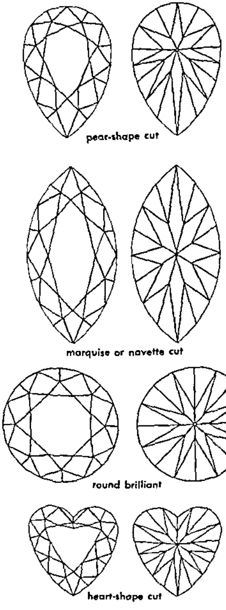


Fig. 2. Examples of brilliant-cut facets. (*Gemological Institute of America*)

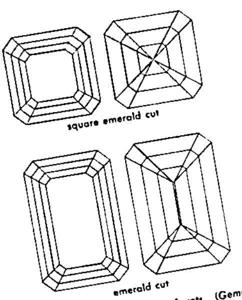


Fig. 3. Examples of step-cut facets. (Gemological Institute of America)

cut stone and its beauty. Rough gem material is so expensive that the natural tendency to retain in the final product every milligram possible often overcomes consideration of potential beauty. Since diamonds of fine quality are valued for their brilliancy and prismatic fire, the need for cutting to precise angles and proportions is appreciated by most diamond cutters, even though many depart somewhat from ideal figures to obtain larger stones that can be sold at lower prices. Since colored stones are valued for their beauty of color, some lapidists believe that any symmetrical cutting is satisfactory. Actually, the lower refractive indices of the important colored stones reduce both their potential brilliancy and their selective absorption, which produces the color. Many colored stones are so poorly cut that the reflective quality of the center portion is reduced, thus making it possible to see the finger of the wearer through the stone. That section of the stone, which is the most important to its ultimate beauty of color, becomes almost colorless, for no appreciable quantity of light is reflected back to the eye; therefore, much more selective absorption is lost.

Cutting styles. Innumerable cutting styles are used to fashion gem stones. The two basic types are those that employ curved surfaces, which are called cabochons, and those with flat surfaces, which are said to be faceted.

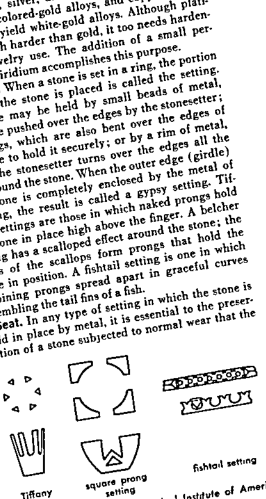
The two basic arrangements of facets are known as the brilliant and the step cut, from which most cutting styles derive. Both are characterized by a large facet (table) topping the crown facets, which slope to the periphery of the stone (girdle), and by a base portion, which slopes from the periphery to a tiny facet (culet) at the lowest point (Fig. 1). The two styles differ in that the facets on step cuts are parallel to the table and the girdle, so that they are all trapeze-shaped, whereas brilliant-cut facets, other than the table and culet, are triangular or kite-shaped. The brilliant style is usually used for stones with round or curved girdle outlines (Fig. 2), and the step cut for those with square, rectangular, or other straight-sided girdle outlines (Fig. 3). The brilliant forms illustrated include the pear shape, marquise, round brilliant, and heart shape. A square step cut and the so-called emerald cut exemplify the step-cut style. See DIAMOND; *see also* CAMEO; INTAGLIO (GEMOLOGY). [R.T.L.]

Gem mounting

Natural gem stones are usually set in rings or other jewelry pieces made of one of the precious metals. Those metals most widely used for jewelry purposes are yellow gold, white gold, platinum, and palladium. Gold alloys are described by the karat system of fineness, in which pure gold is 24 karat (k). The number stamped inside the shank of a ring or on the back of a jewelry piece indicates the number of parts of gold based on a total of 24. Since gold is much too soft to be practical for use as a jewelry metal, other materials are added to harden it. The most popular alloys are 14k and 18k. Copper, silver, and zinc are commonly used to produce colored-gold alloys, and copper, nickel, and zinc to yield white-gold alloys. Although platinum is much harder than gold, it too needs hardening for jewelry use. The addition of a small percentage of iridium accomplishes this purpose.

Setting. When a stone is set in a ring, the portion in which the stone is placed is called the setting. The stone may be held by small beads of metal, which are pushed over the edges by the stonesetter; by prongs, which are also bent over the edges of the stone to hold it securely; or by a rim of metal, which the stonesetter turns over the edges all the way around the stone. When the outer edge (girdle) of a stone is completely enclosed by the metal of the ring, the result is called a gypsy setting. Tiffany settings are those in which naked prongs hold the stone in place high above the finger. A belcher setting has a scalloped effect around the stone; the edges of the scallops form prongs that hold the stone in position. A fishtail setting is one in which adjoining prongs spread apart in graceful curves resembling the tail fins of a fish.

Seat. In any type of setting in which the stone is held in place by metal, it is essential to the preservation of a stone subjected to normal wear that the



Gem-stone settings. (Gemological Institute of America)

"seat" into which it is placed be even. The seat is usually cut into the metal of the setting, after which the stone is pushed into position and the metal beads, prongs, or metal bezel are forced down around the edge of the stone. Ring settings and the various types of stone settings employed in other types of jewelry set with natural stones thus depend on metal pressing on both top and bottom of the outer edge of the stone. If the seat is uneven, the pressure on the stone is applied unevenly, and even a diamond may cleave or fracture when subjected to the light blows that it is certain to receive in the course of normal wear. [R.T.L.]

Gemini

The Twins, in astronomy, is a winter zodiacal constellation. Gemini is the third sign of the Zodiac. It is conspicuous, containing first-, second-, and third-magnitude stars. These stars form a rough quadrilateral figure. The constellation is pictured as the figures of the twin heroes Castor and Pollux, with the two brightest stars of the same names representing the heroes' heads. Pollux, slightly brighter than Castor, is a navigational star. The Sun is in this constellation at the time of the summer solstice. The southwest corner of Gemini, being in the Milky Way, contains star fields which are among the finest in the sky *See* CONSTELLATION. [C.S.Y.]

Gemology

The science of those minerals and other materials which possess sufficient beauty and durability to make them desirable as gem stones. It is concerned with the identification, grading, evaluation, fashioning, and other aspects of gem stones. *See* GEM.

For many years, the little systematic attention given the subject was divided between jewelers and mineralogists. Mineralogy itself gained the status of an exact science only in the last century. Gemology, which has developed within the last several decades, was given impetus by the accelerating developments in synthetics, imitations, various means of color alteration practiced to defraud, and a growing need for methods of detecting these products. *See* GEM, MANUFACTURED. [R.T.L.]

Gene

The basic unit in inheritance. There is no general agreement as to the exact usage of the term, since several criteria that have been used for its definition have recently been shown not to be equivalent.

The nature of this difficulty can be indicated most easily after a description of the earlier position. The facts of Mendelian inheritance indicate the presence of discrete hereditary units that replicate at each cell division, producing remarkably exact copies of themselves, and that in some highly specific way determine the characteristics of the individuals that bear them. The evidence also shows that each of these units may at times mutate, to give a new equally stable unit, which has more or less similar but not identical effects on the characters of its bearers. Each unit can be shown to occupy a specific locus in a chromosome, and the new units (alleles) to which it gives rise occupy the same locus. *See* MENDELISM.

These hereditary units are the genes, and the criteria for the recognition that certain genes are alleles have been that they (1) arise from one another by a single mutative step, (2) have similar effects on the characters of the organism, and (3) occupy the same locus in the chromosome. It has long been known that there were a few cases where these criteria did not give consistent results, but these were explained by special hypotheses in the individual cases. In recent years, however, such cases have been found to be so numerous that they appear to be the rule rather than the exception. For a more complete discussion *see* ALLELE; GENE ACTION; MUTATION; RECOMBINATION, GENETIC.

Some authorities use the term gene to indicate the smallest unit of recombination. This is a logical procedure, but a somewhat inconvenient one operationally, since there is no method of determining when the limit of divisibility has been reached. Other authorities use the term to designate an area in a chromosome made up of subunits that are closely related in their action and that must be present in an unbroken unit to give their characteristic effect. This definition has the disadvantage of being rather indefinite, since intermediate conditions are known.

One result of the varying definitions is a tendency to discard the word gene entirely and to substitute the terms muton (mutational unit), cistron (unit of biochemical activity), and recon (unit of genetic recombination) for it.

The chemical nature of the genetic material is now being actively studied (*see* CHROMOSOME). It is probable that with increasing knowledge of the nature and properties of deoxyribonucleic acid (DNA) it will become possible to reach a more generally acceptable solution to the problems of terminology outlined above. *See* DEOXYRIBONUCLEIC ACID; NUCLEIC ACID. [A.H.ST.]

Gene action

The functioning of genes (hereditary units) in determining the structural and functional characteristics of an individual, that is, its phenotype. Gene action is studied by two somewhat different, but complementary, approaches: (1) the analysis of changes which occur in the phenotype when a gene mutates, or is changed in dosage, or position relative to other genes; this is frequently called the study of phenogenetics; and (2) the more direct approach which attempts to determine the actual means by which genes exert their control over metabolism. The more direct approach is best described as the study of primary gene action, although it includes the study of the interaction of the primary or secondary products of gene action. *See* GENE.

The study of gene action is equal in importance to the study of gene replication, and as such consti-

tutes one of the most fundamental fields of investigation in biology. An understanding of gene action is basic to the understanding of metabolism and development in plants and animals, and of disease, both inborn and secondarily contracted.

Phenotypic effects of gene mutation. An almost infinite variety of phenotypic changes can be expected from the mutation of an organism's genes. For descriptive purposes, mutant genes can be grouped into a number of categories based on the effect they have on the phenotype. A few examples are given here which serve to illustrate the wide spectrum of effects expected from mutation.

It is important to recognize that any mutant gene may fit all or most of the following categories. The category in which it is placed depends to a very large extent on the viewpoint of the observer.

Lethal mutations probably occur more frequently than any other kind. These cause the death of the organism at as early a stage as the egg. Death may also occur at later periods, usually at a rather specific time in the development of the organism. Lethals may be either recessive or dominant. *See* LETHAL GENE.

Semilethals and sublethals are mutant genes which have less than a completely lethal effect. The organism carrying them is definitely impaired in its ability to cope with the environment, and may have a shorter life span than the normal.

Minor detrimentals are a class of mutants not easily or necessarily distinguishable from those noted above. These mutant genes do not have a drastic effect on the phenotype, but they impair the functioning of the organism sufficiently so that it may not compete successfully with other genotypes.

Neutral mutations show no marked selective advantage or disadvantage.

Beneficial mutant genes may be more efficient than their wild-type alleles in providing for a phenotype with a distinct selective advantage.

Morphological, mutant genes cause changes in appearance or structure. In microorganisms, changes in growth habit may produce differences in the appearance of a colony. These are also called morphological mutants.

Biochemical mutants are those in which a definite step in biosynthesis can be shown to be brought about in some way by the mutation. Actually, all mutations cause basic biochemical changes.

Sterility mutants are incapable of producing functional gametes, or produce gametes of restricted ability to fertilize or be fertilized. These mutations may affect either plants or animals. In plants, certain types of sterility mutants are important in preventing self- or cross-pollination.

Fecundity mutations permit the production of gametes, but at a lower level than the normal alleles.

Genes as units of action. Genes are ordinarily considered as units of action or function. Hereditary characteristics are not inherited as such, but the genes which determine them are inherited. Since the hereditary traits appear to be inherited as units, the genes controlling them are most logically conceived as units of function. There are certain difficulties with this interpretation, because the physical gene or crossover unit, the recon, may not necessarily correspond exactly with the functional gene, which may consist of a number of recons. *See* PSEUDOALLELES.

It has been suggested that the name cistron be given to the functional gene, or functional unit. This unit is recognized only when homologous chromosome segments carrying the units are together, as in the cells of diploids. Two linked genes *a* and *b* which are recessive and have similar mutant effects (as do alleles) and are separable by crossing over (that is, are different physical units) may be spatially related in one of two ways in the double heterozygote, as follows:

$$\frac{a+}{+b} \quad \textit{trans} \tag{1}$$

$$\frac{ab}{++} \quad \textit{cis} \tag{2}$$

If the phenotype is mutant (like or similar to that expected for either mutant gene homozygous, that is, *aa* or *bb*) when the mutant genes are in the trans arrangement, and normal (wild-type) when the genes are together in the cis condition, the two recombinational units are said to be part of the same functional unit or cistron. It should be noted that if the two genes were not related as part of a functional unit, both the cis and trans arrangements should give the same normal phenotype.

Primary gene action. The actual effects of gene action are recognized for the most part by noting the effects of gene mutation on the phenotype, but the final phenotypic effect must be considered to be far removed from the initial action of the gene itself. By studying the phenotype, however, particularly in the lower organisms such as bacteria and fungi, it has become evident that there is a close relationship between genes and enzymes. The relationship may be very direct. Each gene in the genome (complete set of hereditary factors) may be responsible for the production of a single enzyme, the one gene–one enzyme hypothesis, which is a form of the one gene–one function hypothesis mentioned previously.

The close connection between the action of genes and the appearance of enzyme activity has been noted in a number of different microorganisms and in plants and animals, including man. Many mutant genes have been described which cause the disappearance or inactivation of specific enzymes. The table on the following page gives a number of examples drawn from a variety of sources to illustrate the widespread occurrence of this phenomenon.

Since the absence of an enzyme causes a reactio to be blocked, and since its absence is inherit the block is called a genetic block.

Inherited conditions in which an enzyme activity is absent or not detected

Organism and condition	Reaction involved	Name of enzyme
Man		
Galactosemia	Galactose-1-phosphate + Uridine diphosphoglucose ⇆ Glucose-1-phosphate + Uridine diphosphogalactose	Galactose-1-phosphate-uridyl transferase
Alkaptonuria	Homogentisic acid → Maleyl acetoacetic acid	Homogentisate oxidase
Phenylketonuria	Phenylalanine + O_2 → Tyrosine	Phenylalanine hydroxylase
Neurospora		
Requires tryptophan	Indole + Serine → Tryptophan	Tryptophan synthetase
Rabbit		
No atropine hydrolysis	Atropine → Tropine + Tropic acid	Atropinesterase
Clover		
Acyanogenesis	Linamerin → Glucose + Acetone + Cyanide	Linamerase
Yeast		
Inability to use galactose	Galactose + ATP ⇆ Galactose-1-phosphate + ADP	Galactokinase
Escherichia coli		
Requires isoleucine + valine	{Ketovaline or Ketoisoleucine} + Amino acid (carried by donor) ⇆ {Valine or Isoleucine} + Keto acid	Transaminase B
Drosophila melanogaster		
Rosy and maroonlike eye mutants	2-Amino-4-hydroxypteridine → Isoxanthopterin	Xanthine dehydrogenase

In addition to causing the complete absence of detectable enzyme activity, gene mutation may also cause a reduction in enzyme activity to a low level, yet above zero, or may even increase the activity above the normal level. Furthermore, mutation may result in the formation of enzymes with activity comparable to wild type, but with different physical characteristics.

An enzyme is a protein macromolecule which is detected usually by its activity for catalyzing a specific reaction or type of reaction. Other methods exist for detecting and identifying macromolecules which may or may not have enzymatic activity. Some of these methods are electrophoresis, antigenic specificity tests, and ultracentrifugation. *See* ENZYME.

Different hemoglobin species have been identified in man by [illegible] of hemoglobin in the protein component, but not in the heme component. Amino acid analyses of the protein components, in which the sequence as well as the identity of the amino acid residues was established, show that the hemoglobin S molecule differs from the A hemoglobin in only one amino [illegible] ence is ascribable to a single gene difference. Hemoglobin C, another abnormal hemoglobin, inherited through a gene allelic to that determining S, has lysine instead of valine replacing the same glutamic acid residue. These examples indicate that genes may determine the specific configuration of macromolecules, such as proteins, by determining the sequence of amino acids in the protein polypeptide chains. *See* HEMOGLOBIN.

By means of the antigen-antibody test, it is possible to show that rather direct relationships exist between genes and antigens. Series of genes, both allelic and nonallelic, which determine the presence of specific types of antigens, have been found in such diverse kinds of animals as man, chickens, cattle, rabbits, and *Paramecium*. *See* ANTIBODY; ANTIGEN; IMMUNOLOGY.

Within any single species of animal, it is usually found that some antigens are inherited through allelic genes. Thus, in man, the antigens A and B on the surface of the red blood corpuscles are inherited through the allelic genes I^A and I^B, respectively. A third allele I^O is inactive in producing any detectable antigen. Persons homozygous for I^A, I^B, or I^O are respectively of blood types A, B, or O. In the heterozygous condition, I^AI^B, both antigens A and B are present, and the individual is of blood type AB. When either I^A or I^B is heterozygous with I^O, only the specific antigen determined by the active allele is present. *See* BLOOD GROUPS.

From observations such as these, it has been concluded that a direct relationship exists between genes and antigens. Each antigen is determined by a specific gene. Genes forming detectable antigens do so independently of other antigen-producing genes, even those allelic to them and present with them in heterozygotes. Exceptions to this apparent independence are discussed below.

In some animals, particularly in cattle and chickens, overlapping specificities may occur between antigens and antibodies. For example, in cattle, one

gene locus designated as B has a series of alleles, each of which produces a different antigen. However, the different antibodies reacting with these antigens do not always show a one-to-one specificity of antigen to antibody. Antigens B_2, B_3, and B_4, produced in the presence of alleles B^2, B^3, and B^4, respectively, react with their corresponding antibodies, as expected; but in addition, B_3 may react with the antibodies for both B_2 and B_4, although B_2 and B_4 antibodies may not react with each other's antigens. It is apparent that three different antigens may be distinguished, even though the specificities overlap. This may be interpreted as meaning that allelic genes form different but related antigens, as indicated by the overlapping specificities of the antibodies with which they react.

By applying the method of ultracentrifugation, it has been demonstrated that there exist in human blood serum at least two hemoglobin-binding globulins designated as haptoglobulins. These migrate electrophoretically with the α_2-globulins of serum, and can be separated by their difference in sedimentation rate in the ultracentrifuge. Two allelic genes are involved in their inheritance. Individuals homozygous for Hp^1 or Hp^2 carry one or the other of the haptoglobulins. See GLOBULIN; ULTRACENTRIFUGE.

In view of findings of the types described above, the generalization can be made that genes determine the configuration and hence the biological specificity of proteins, and perhaps of other macromolecules in the cell. These then control the metabolic activities of the organism by acting as enzymes and antigens. When a gene mutates to another allele, it produces an altered macromolecule which may or may not perform the original function. If it has activity, it may carry out its function with a greater or lesser degree of efficiency than the original product of the standard allele.

The question remains to be answered of just how direct the relationship is between genes and the specific molecules manifest to the observer as enzymes, antigens, and the like. It is highly probable that these substances are not the primary gene products. Rather, as proteins, they are formed in the presence and under the control of cytoplasmic ribonucleic acid (RNA) which has been shown to be involved directly in the synthesis of protein. The RNA may receive its specificity from the gene deoxyribonucleic acid (DNA). Hence, it may be postulated that RNA is closer to being the primary gene product and that the sequential relationship DNA $\rightarrow$ RNA $\rightarrow$ specific active protein $\rightarrow$ phenotype may closely approximate the actual situation in organisms. Indeed, the analysis of the relationship between DNA, RNA, and protein may well be the most direct approach to the study of gene action, and should give the most fruitful approach to most of the questions and problems discussed here. See CYTOCHEMISTRY, NUCLEIC ACID.

Gene dosage. Under certain conditions, an increase in the number of times a particular gene is present has a direct quantitative effect on one or more aspects of the phenotype. This effect is considered to be a manifestation of quantitative gene activity or gene dosage. It means that the gene does something, and that the higher the dose with which it is present, the greater the physiological or chemical end result. The best examples of this are found in the genetics of polyploid plants (see CHROMOSOME). In these plants, it is possible to increase the dosage of a particular allele from zero to four or more, depending on the degree of ploidy. When this is done, for example, with certain genes which control the production of the flower pigments of the *Dahlia*, there is a demonstrable increase in the amount of pigment in the petals. Also, in maize (Indian corn), increasing the dose of the gene Y as follows: $yyy \rightarrow Yyy \rightarrow YYy \rightarrow YYY$ in the triploid endosperm of the kernels results in the following relative concentrations of carotenoids with vitamin-A activity: 0.05, 2.25, 5.00, and 7.50, respectively. Thus, a single dose of Y determines the presence of approximately 2.50 units of vitamin-A activity. The allele y apparently has no activity for this phenotype, and therefore increasing its dosage has no effect. See POLYPLOIDY; VITAMIN A.

An increase in the dose of an active allele of a gene does not always cause an increase in the manifestation of a particular aspect of the phenotype. It may have quite the opposite effect, and cause a decrease. In the *Dahlia*, an increase in the dose of the gene A which determines the presence of anthocyanin pigments causes an increase in the amounts of these pigments, but a concomitant decrease occurs in the amount of a related pigment, apeginin. Thus, the relationship between dose and effect can be either direct or inverse.

Interactions between alleles. Some heterozygous combinations of mutant alleles do not produce the phenotype expected after observing the phenotypes of the homozygotes. This is defined as a manifestation of allelic interaction. It is in contrast to those situations in which one allele is dominant over the other, so that the phenotype of the heterozygote is very similar or identical to that of the homozygous dominant. Also, it is different from those situations in which the two alleles show an additive effect, and the phenotype of the heterozygote is intermediate between those expected from the homozygotes See DOMINANCE, RECESSIVENESS, BLENDING.

Antimorphic interaction. Mutant alleles which show a positive but weak effect toward producing the wild-type phenotype when homozygous, may show an even weaker effect when present together in heterozygotes.

$$\text{Normal phenotype} > \frac{a^1}{a^1} > \frac{a^2}{a^2} > \frac{a^1}{a^2}$$

Occasionally, one allele may act as an amorph when homozygous (no type effect), but show dominance over other genes allelic to it so that

$$\text{Normal phenotype} > \frac{a^1}{a^1} > \frac{a^0}{a^0} = \frac{a^1}{a^0}$$

a^1 is an active allele and a^0 is an

Alleles of this type are referred to as antimorphs, since they seem to oppose the active production of the specific phenotypic effect given by other alleles in the series. The result is a subtractive rather than an additive effect in combinations. It suggests that a competition or inhibitory interaction between the gene products, or even the genes themselves, may be in operation.

Antigens and other macromolecules in heterozygotes. Animals heterozygous for two alleles, each producing a distinctive antigen, ordinarily produce both antigens. However, when the quantitative amounts or strengths of the antigens produced by the heterozygotes have been determined, it has been found many times that the values for the heterozygotes are in fact lower. Thus, there is an interaction at some level, like that described for antimorphs, which results in a lowered production of each antigen in the heterozygotes.

A phenomenon which is perhaps related is found in heterozygous humans who carry the allelic genes for both normal A hemoglobin and sickle-cell S hemoglobin. Instead of the 1:1 ratio that would be expected if each gene acted completely independently, these individuals have only 25–45% S hemoglobin with the remainder A hemoglobin.

An increasing number of cases is being discovered in which a new molecular species is produced in hybrids or heterozygotes, but not in the

duces the antigen C, and the species *Columba livia* the antigen C′. The hybrids from these two parents produce an additional antigen not found in either parent and called the hybrid substance. A similar hybrid substance has been found in the rabbit and results from the combination of the allelic genes Hg^A and Hg^D in heterozygotes. In these, both antigens A and D are produced, but an additional one, I, also appears which cannot be detected in the homozygotes, Hg^AHg^A and Hg^DHg^D.

Humans presumably heterozygous for the allelic genes Hp^1 and Hp^2, which produce the two distinctly different haptoglobulins mentioned above, have been found to produce a third haptoglobulin. This can be distinguished from the haptoglobulins found in the homozygotes by differences in sedimentation rate and electrophoretic mobility.

The development of new methods for analyzing the chemistry of the phenotype and for distinguishing subtle differences in macromolecules makes it possible to study interactions in diploids and or-

hybrid vigor. *See* HETEROSIS.

Gene interaction. The final

mary products, but the primary products, and the products of their activity, that is, enzymes and other macromolecules, interact at the level of extragenic metabolism to give the final phenotype. Thus, there is really no one gene determining the shape of an organ, or even the production of a certain pigment. These end-products are determined by many genes acting together through their respective immediate and then succeeding interrelated products. This is true even though the mutation of only a single one of the genes may cause a characteristic unit change in the phenotype. The manifestation of these interactions, as determined initially by the results from breeding experiments and in a few cases from biochemical analysis, is called gene interaction. This term does not necessarily imply that the genes themselves interact. In general, the term is applied to apparent interactions between genes. A few examples of certain types of gene interaction are given in the following discussion.

Complementary genes. These are nonallelic genes so directly involved in the formation of the same end-product, or phenotype, that the mutation of any one of them to an inactive state will result in no end-product or type effect. All must be active at once. Figure 1 illustrates the relationship between a series of biosynthetic reactions in *Neurospora* leading up to the formation of nicotinic acid. Obviously, the failure of any one of the indicated genes to act will produce a requirement for nicotinic acid. This vitamin is not required to be present in the growth medium of the wild-type strain.

Epistasis. Epistasis is a general term applied to cases in which one gene masks the effects of other genes that may be present. For example, each gene controlling a step in the reaction sequence shown in Fig. 1 is potentially epistatic to those later in the sequence. Epistatic gene action is particularly noted in the pigmentation of animals. The white Leghorn fowl has unpigmented plumage because of the presence of a dominant gene which is epistatic to all color genes which modify the pigmentation of the feathers in its absence. In *Drosophila melanogaster*, at least 40 different genes are known to modify the pigmentation of the eyes. They produce hues and shades of brown to brilliant red of several degrees of intensity. However, in the presence of the white gene *w* homozygous, the eyes are colorless, regardless of the residual genotype.

Suppressor genes. Suppressor genes are those which cause a wild-type or normal phenotype despite the presence of nonallelic mutant genes. In a sense, these genes are also epistatic, since they mask mutant effects. The mutant gene vermilion *v* in *D. melanogaster* causes an eye to develop without the brown pigment component which is present in the eye of the wild type. It has been shown that failure to develop the brown pigment in *vv* flies is the result of a genetic block intervening between tryptophan and formylkynurenine, two precursors of the brown pigment. Another gene, the suppressor of vermilion (*su-v*), gives a wild phenotype when present homozygous with *v* (*vv, su-v/su-v*). Biochemical analysis of the suppressed mutant

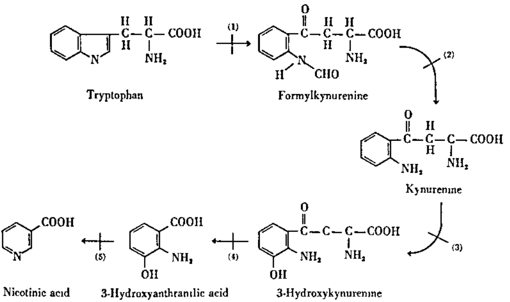


Fig. 1. Nicotinic acid biosynthesis.

shows that it can produce formylkynurenine from tryptophan. Hence, the suppressor has relieved the block in metabolism, even though the mutant gene *v* is present.

Suppressor genes have, in at least one case, been shown conclusively to cause the appearance of an enzyme genically determined to be absent by another gene. The mutants of the fungus *Neurospora crassa*, which cannot synthesize tryptophan from indole and serine, because of the absence of the enzyme tryptophan synthetase brought about by the mutant gene *td*, have been most useful in demonstrating this. When the mutant gene is present with a suppressor gene, *su-td* (*td, su-td*), the mold grows in the absence of tryptophan, and an active enzyme can be demonstrated in it which cannot be distinguished from that produced in the wild type. There is no evidence that the suppressor gene has anything directly to do with the production of the enzyme. Rather, the indications are that it modifies the action of the *td* gene in some way so that an active enzyme is produced. This modifying activity may include such things as reducing the amount of a substance which is inhibiting the activity of the enzyme produced by the *td* gene.

Suppressor genes have many peculiarities. They may have no apparent effect on the phenotype other than their suppression of the effects of a mutant gene. Certain of them are known to suppress the action of a number of unrelated and nonallelic mutant genes. Allelic or pseudoallelic genes may respond differently to suppressors. Only one member of an allelic series may be suppressed by a specific suppressor. Other alleles in the same series may be suppressed by still other suppressor genes.

Modifiers. With respect to its control over the phenotype, every gene has its final effect influenced by the action of other genes. In a sense, all genes within the genetic system of the organism modify one another's activities directly or indirectly. The examples given above of complementary and suppressor gene activity serve to illustrate a drastic type of modifier effect in which one gene acts as a modifier to erase another's effect completely. Yet many modifiers have only minor effects on the actions of other genes. For example, they may enhance the mutant effect of another gene (enhancers), or they may dilute a pigment determined to be present by another gene (dilution genes). Their effects may even be so slight that they are recognized only when groups of them act together. In this event, the particular character they affect may be inherited in a quantitative manner. Minor modifiers of this type are sometimes referred to as polygenes to distinguish them from genes with large effects, the oligogenes. It is to be recognized, however, that the minor modifiers are minor only in the sense that they or their usual variants have small effects on the phenotype when placed in different combinations by recombination. It is not to be assumed that they are minor in function. A major mutation of any one of them might cause a major effect on the phenotype, including even a drastic lethal effect.

Quantitative inheritance. Many characteristics, such as height, weight, size, and intelligence, show a quantitative and continuous type of variation. This is opposed to the discontinuous type of variation in which only one or two pairs of allelic genes segregate in a cross and the phenotypic classes can be readily recognized as different, with no intergrades.

Perhaps quantitative inheritance can best be illustrated and understood by considering the results of a breeding experiment with Indian corn of long-eared and short-eared varieties. A mixed population will show a binomial distribution with respect to ear length. A few with very short

very long ears will be represented, and the others will be various intergrades between the extremes. If one inbreeds the shortest-eared and the longest-eared of these to obtain approximately true-breeding short-eared and long-eared, and then crosses these together, one will not obtain types like the parents in the F_1. Rather, the F_1 offspring will be intermediate in ear length Plants with long ears and short ears can be obtained by inbreeding the F_1 plants. The distribution in regard to ear lengths expected in the F_2 is illustrated in Fig. 2.

The usual interpretations of these results is that the character ear length is, as might be expected, the resultant of a great many genes, each of which makes a minor contribution to ear length. In general their effects are additive, and taken together, they show a cumulative effect. If it is assumed that *a series of nonallelic genes* T_1, T_2, T_3, T_4, T_5 *act* additively to produce long ears, whereas their alleles t_1, t_2, t_3, t_4, t_5 act to produce short ears, then plants homozygous for $T_1, \ldots, T_5$ will have the longest ears, and those homozygous for $t_1, \ldots, t_5$ have the shortest ears. If these homozygous extremes are crossed, the F_1 offspring will all be heterozygous and intermediate in height, provided T genes are not completely dominant to their t alleles. Crossing the heterozygotes will then give offspring with all the possible combinations of these,

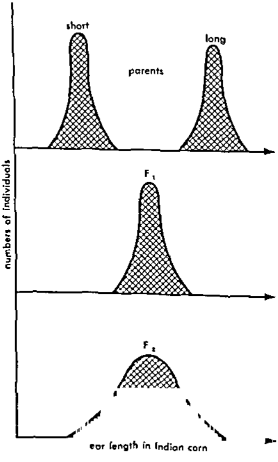


Fig. 2. Quantitative Inheritance.

including the homozygous extremes. A binomial distribution with respect to height is expected on the basis of the Mendelian laws. This interpretation of quantitative inheritance is known as the multiple factor hypothesis. It fits most of the known data from breeding experiments in which quantitative characters are considered.

Two individuals which show a large quantitative difference usually produce an intermediate F_1 generation. Occasionally, however, when the F_1 is inbred, the F_2 or F_3 generations may include individuals which are more extreme than either of the original parents. This is known as transgressive variation. It is easily explained by assuming that the original parents, while perhaps homozygous, were not homozygous for all of the genes giving the extremes. To use the example given in the preceding paragraph, they might have been homozygous for T_1, T_2, T_3, t_4, t_5 and t_1, t_2, t_3, T_4, T_5, respectively. It is evident that hybrids produced from these will be expected to give some complete homozygotes for the T or t genes, respectively, when crossed. These will, of course, be more extreme than the parents.

Gene and environment. The phenotype which develops in any organism is the resultant of two guiding influences, the genotype and the environment (*see* GENETICS). Hence, the environment must be considered in any analysis of gene action, if it is desired to arrive at an understanding of how genes act toward the production of the phenotype. Practically, this is best done by keeping the environment as constant as possible while making studies of gene action. However, much also can be learned by varying the environmental conditions and keeping the genotype as constant as possible.

The environment of genes is obviously a complex one. For convenience, two areas can be defined: (1) that immediately around the genes, the intracellular environment of the rest of the cell; and (2) the extracellular and extraorganismal environment. The intracellular environment can be changed by the mutation of other genes, which may then modify the action of a gene under study. Gene interaction is thus seen to be in part an aspect of the study of the internal environment of the cell.

Extracellular environmental factors, such as light and heat, may also influence the action of genes greatly. For example, in *Neurospora* there are some mutant strains which at one temperature have absolute growth factor requirements, but not at another temperature 10° removed. At the latter temperature they will grow as well as the wild type on the minimal medium with no supplement. In *Drosophila*, an eye color mutant gene, blood, w^{bl}, permits the production of the eye pigments only at low temperatures. At temperatures above 25°C little or no pigment is produced in the eyes. Even the types of antigens produced by an organism may be influenced by the environment. This has been demonstrated in *Paramecium*. In this protozoan, transformations in antigenic type can be brought about by a change in temperature or by several

other environmental modifications. The transformation need not be permanent, for if the original environment is restored, the original antigenic type will reappear.

Changes in phenotype which occur against a constant genetic background are in reality responses to the environment by the extragenic part of the living system. The genes themselves are not changed, as can be readily demonstrated by changing the environment back to the original condition, or by breeding the individual and showing that the offspring inherit the original parental genotype.

Phenocopies are particular types of environmentally induced changes which mimic in appearance mutant types produced by mutant genes. For example, by means of heat shock or chemical treatment of larvae of wild-type *D. melanogaster*, a number of different types of aberrant development may be induced that will result in the formation of adults with characteristics which superficially resemble those possessed by certain mutants. These environmentally induced modifications are not inherited, however, whereas those resulting from mutant gene action are inherited. See EXPRESSIVITY, GENE; PENETRANCE, GENE; PHENOCOPIES. [R.P.W.]

GENE ACTION IN DEVELOPMENT

Genes as indispensable units. Changes or losses of genes through mutation lead in a high percentage of cases to a breakdown of development. This follows most clearly from studies on chromosomal deficiencies. Genotypes of the fruit fly *Drosophila*, which lack only one out of several thousands of loci visible in their giant salivary chromosomes, are, as a rule, not able to develop beyond a definite stage. That is, to say, among all the other loci not affected by the mutation, none is capable of replacing the function of the missing one. Each gene is therefore characterized by its locus-specific and indispensable action. Consequently, losses of different genes cause developmental abnormalities different in nature and degree. Normal development is a highly sensitive process dependent on the harmoniously integrated function of thousands of genic units.

Phase specificity of action. Not all of the genes of an organism are needed at the beginning of development; but the more development and differentiation proceed the more genes have to contribute their specific functions. Thus, the gene-conditioned products come into action along the time axis of development in a stepwise and phase-specific manner. Certain genes begin to function during the formation of the egg cell in the ovary. They thereby influence prior to fertilization the developmental fate of the next generation. This phenomenon is known as maternal effect, or predetermination. Some paternal genes introduced by the sperm nucleus can likewise exert their specific effects almost immediately after fertilization. Most of the genes, however, manifest themselves later on. In all higher organisms, there are limited periods during which especially numerous and various gene products are needed. In insects, the onset of metamorphosis is such a sensitive phase. In birds and mammals, the time of hatching or the days shortly before or after birth, respectively, are developmental phases which require the normal functions of a great many genes. Mutations therefore lead, especially often, to lethal effects occurring within these periods. The sensitive phases are moreover susceptible to various external influences, such as extreme temperatures, poisonous chemicals, ionizing radiations, and deficiencies in nutritive substances. These environmental agents, when acting during a sensitive phase, often cause developmental changes and abnormalities that correspond exactly with effects of known hereditary factors. Such an imitation of the manifestation of a gene, produced by an external factor in an unmutated genotype, is called a phenocopy.

Special studies have shown how phase specificity can be due to biochemical effects of mutations. Such an example is illustrated in Fig. 3. The gene lethal-meander *lme* of *Drosophila* causes a standstill of growth during the third larval instar. The dwarfed larvae never enter metamorphosis and eventually die because in these genotypes, the supply of essential amino acids and peptides needed for protein synthesis is blocked.

Cell and organ specificity. Individual genes, though present in all developmental systems of an organism, may affect the hereditary characters

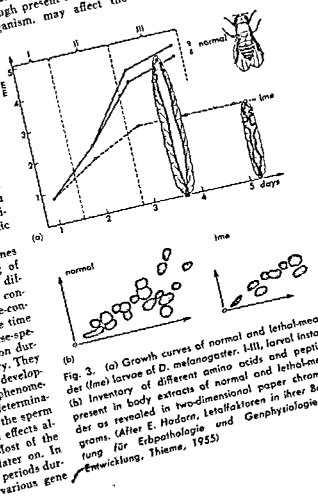


Fig. 3. (a) Growth curves of normal and lethal-meander (*lme*) larvae of *D. melanogaster*. I-III, larval instars. (b) Inventory of different amino acids and peptides present in body extracts of normal and lethal-meander as revealed in two-dimensional paper chromatograms. (After E. Hadorn, Letalfaktoren in ihrer Bedeutung für Erbpathologie und Genphysiologie der Entwicklung, Thieme, 1955)

(phenes) of only a selected group of cells or organs. In mammals, for instance, genes are known whose action appears respectively confined to red blood cells, to rods and cones in the retina, to cartilage, to certain specialized cells of the nervous system, or to hormone-producing cells of the pituitary. Other mutations interfere with the formation of selected groups of whole organs. Within the skeleton, either the development of the extremities alone might be blocked (Fig. 4*a*), or the limbs may be formed normally whereas the vertebral column becomes shortened and severely malformed (Fig. 4*b*). In man, a certain dominant gene acts in a surprisingly restricted manner by causing a shortening only of the middle phalanx of the index finger and of the second toe (Fig. 5). Cell specificity of genic action must be due to those biochemical differences on which any and all developmental differentiation is based. Since different cellular systems and organs require for their specific makeup different gene-conditioned products, they become diversely or not at all affected by a particular mutation. This mechanism leads to the phenomenon of pleiotropy.

Primary versus secondary effects. Hereditary characters (phenes) which are manifest in cellular systems are either conditioned directly by the genes acting within the character-forming cells, or

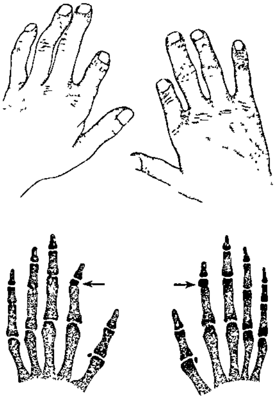


Fig 5. Localized action of a gene in man. Arrows in the x-ray pictures point to the affected element. (*After O. L. Mohr and Ch. Wriedt, from E. Hadorn, Letalfaktoren in ihrer Bedeutung für Erbpathologie und Genphysiologie der Entwicklung, Thieme, 1955*)

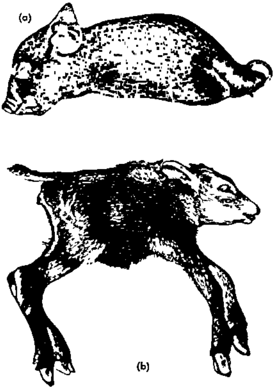


Fig. 4. Organ-specific action of lethal factors. (a) Streamlined piglet (*after L. E. Johnson*); (b) malformed calf (*after O. L. Mohr and Ch. Wriedt*). (*From E. Hadorn, Letalfaktoren in ihrer Bedeutung für Erbpathologie und Genphysiologie der Entwicklung, Thieme, 1955*)

they are the secondary consequences of genic actions taking place primarily in other cells of the same organism. The first category of intrinsically caused characters has been called autophenes, while the extrinsically conditioned characters behave as allophenes. Secondary developmental effects of mutations will arise whenever the altered hereditary constitution affects primarily the cells and, consequently, the functioning of embryonic inductors or hormone-producing organs. All other cellular systems whose differentiation, growth, and function depend on the changed inductor or endocrine gland will manifest secondary characters (allophenes).

The gene dwarf *dw* of the house mouse interferes specifically with the development of a definite cell type in the pituitary which produces in normal animals the growth- and the gonad-stimulating hormones. Dwarfism, as well as undeveloped gonads and sterility, is a secondary characteristic of this mutant. These abnormal allophenes disappear in dwarf mice which receive daily injections of genetically normal pituitary cells. The primarily conditioned damage in the endocrine cells cannot be cured, however, by this method. It is the aim of developmental genetics to find out by appropriate experiments in what causal relationships the hereditary characters are formed, and how and where genes act primarily.

Autonomous versus nonautonomous development. When implanted into the body of a different strain, cell systems that are affected only secondarily by extrinsic gene effects develop characters corresponding to the genetic constitution of the host. Only in rare cases, on the other hand, is it possible to overcome the influence of a gene with respect to its primarily conditioned character. Numerous transplantation and grafting experiments, using various animals and plants, have shown that transplanted tissues and organ rudiments develop autonomously. They do not manifest the hereditary traits of the host organism but develop those characters which correspond to their intrinsic genetic constitution.

If, for instance, pieces of skin of newly hatched chicks which differ genetically with respect to color and structure of feathers are reciprocally exchanged, a fully autonomous development is observed (Fig. 6). The hereditary characters of the

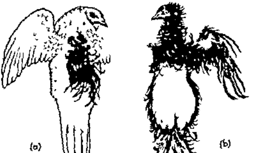


Fig. 6. Autonomous behavior of skin transplants (a) White Leghorn host with implant from a black Frizzle chick. (b) Normal feathering and unchanged color in an implant from a White Leghorn donor developing in a black Frizzle host (After experiments of W. Landauer and S. D. Aberle, from E. Hadorn, Letalfaktoren in ihrer Bedeutung für Erbpathologie und Genphysiologie der Entwicklung, Thieme, 1955)

implants are not influenced by the host. Moreover, cellular autonomy is responsible for the fact that, in higher organisms, tissue transplants from foreign donors seldom survive permanently. Apparently, it is difficult to introduce in a cell such foreign gene-conditioned substances which are able to replace the intrinsically formed gene products or to compete successfully with them.

In a few instances, however, nonautonomous development has been demonstrated. Eye rudiments of the *Drosophila* mutant *vermilion* become wild-type in color when implanted into normal hosts. Here, kynurenine, a pigment precursor which vermilion cells cannot form, is furnished by the host to the implanted cells. Another *Drosophila* mutant, rosy, is deficient in the enzyme xanthine dehydrogenase. Consequently, the pteridine and purine metabolism becomes affected. Normal development could be initiated in mutant organs by growing them in hosts which are capable of forming the gene-conditioned enzymes. Nonautonomous mutants are most valuable research objects on which biochemical and developmental action of genes can be analyzed. *See* MATERNAL INFLUENCE; PLEIOTROPISM. [E.H.]

Bibliography: E. Hadorn, Patterns of biochemical and developmental pleiotropy, *Cold Spring Harbor Symposia Quant. Biol.* 21:363, 1956; E. Hadorn, Role of genes in developmental processes in W. D. McElroy and B. Glass (eds.), *The Chemical Basis of Development*, 1958; W. D. McElroy and B. Glass (eds.), *The Chemical Basis of Heredity*, 1957; C. Stern, Gene action, in B. H. Willier, P. A. Weiss and V. L. Hamburger (eds.), *Analysis of Development*, 1955; C. H. Waddington, *The Strategy of the Genes*, 1957; R. P. Wagner and H. K. Mitchell, *Genetics and Metabolism*, 1955.

Generator, electric

Any machine by which mechanical power is transformed into electric power. Generators fall into two main groups, alternating-current (ac) and direct-current (dc). They may be further classified by their source of mechanical power, called the prime mover. Generators are usually driven by steam turbines, hydraulic turbines, engines, gas turbines, or motors. Small generators are sometimes powered from windmills, or through gears, belts, friction, or direct drive from parts of vehicles or other machines. *See* PRIME MOVER.

Theory of operation. The theory of operation of most electric generators is based upon Faraday's law. When the number of lines (maxwells) of magnetic flux linking a coil of wire is caused to change, an electromotive force proportional to the product of the number of turns times the rate of change of flux is generated in the coil. The instantaneous induced voltage is

$$e = -n(d\phi/dt)10^{-8} \quad \text{volt} \qquad (1)$$

where n is the number of turns, ϕ is the flux in maxwells, and t is time in seconds. The minus sign indicates that the induced voltage opposes the effect which produced it. Voltage is induced in the

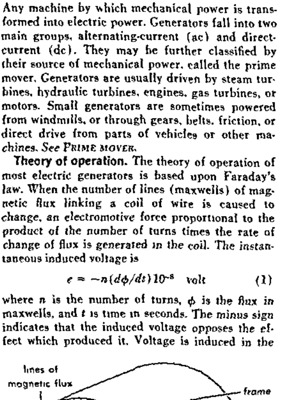


Fig. 1. Elementary generator.

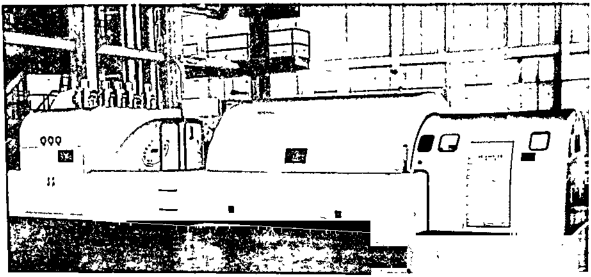


Fig. 2. Turbogenerator unit with direct-connected exciter, rated 22,000 kw. (*Allis-Chalmers*)

windings of a generator by mechanically driving one member relative to the other, thereby causing the magnetic flux linking one set of coils, called the armature windings, to vary, pulsate or alternate. The magnetic flux may originate from a permanent magnet, a dc field winding, or from an ac source. Figure 1 shows an elementary generator having a stationary field and a single, rotating, armature coil. It is apparent that the magnetic flux threading the coil reverses direction twice per revolution, thereby generating one cycle of voltage in the armature coil for each revolution. If this variation in flux is expressed as a function of time, the voltage generated in the coil would be given by Eq. (1). For example, let the flux vary as given by

$$\phi = \Phi_m \cos 2\pi ft \qquad (2)$$

where Φ_m is the maximum flux, f is the frequency, and t is time. By differentiating Eq. (2) with respect to time, and substituting in Eq. (1), the instantaneous voltage in the coil is found to be

$$e = 2\pi fn\Phi_m \sin 2\pi ft \times 10^{-8} \qquad (3)$$

This ac voltage can be taken from the armature by brushes on the slip rings shown. If the coil terminals were brought to a 2-segment commutator instead of slip rings, a pulsating dc voltage would appear at the brushes. See ALTERNATING-CURRENT GENERATOR; DIRECT-CURRENT GENERATOR.

Construction. In practice, permanent-magnet fields are used only in small generators. Large generators, except induction generators, are equipped with dc field windings. The field coils are wound on the stators of most dc generators to permit mounting the armature coils and commutator on the rotor. On ac generators, the field coils are normally located on the rotors. Field coils require only low voltage and power, and only two lead wires. They are more easily insulated and supported against rotational forces and are better suited to sliding contacts than the relatively high-voltage armature windings, which often have six leads brought out.

Any part of the magnetic circuit not subject to changing flux may be of solid steel. This includes the field poles of dc machines and portions or all of the rotating field structure of some ac generators. In machines with small air gaps the poles are frequently of laminated steel, even though their flux may be substantially constant. Laminations help minimize pole-face losses arising from tooth-frequency pulsations. The armature core is almost always composed of thin sheets of high-grade electrical steel to reduce core loss (*see* CORE LOSS).

The windings are insulated from the magnetic structure, and are either embedded in slots distributed around the periphery or mounted to encircle the field poles. The terminals from the stator windings and from the brush holders are usually brought to a convenient terminal block for external wiring connections. *See* WINDINGS (ELECTRIC MACHINERY).

Turbogenerators. Generators driven by steam or gas turbines are sometimes called turbogenerators. Although in small sizes these may be gear-driven, and some may be dc generators, the term turbogenerator generally means an ac generator driven directly from the shaft of a steam turbine. A typical turbogenerator unit is shown in Fig. 2.

In order to achieve maximum efficiency, the steam turbine must operate at high speed. Consequently, direct-connected turbogenerators are seldom built to operate below 1500 rpm. To minimize windage loss and to keep rotational stresses down to a safe level, turbogenerator rotors are usually long and slender, in some cases 5–6 times the diameter in length of active iron. Long rotors operate above their first critical speed, and in some cases near or above their second, thereby introducing mechanical problems in balancing and resonance (*see* MECHANICAL VIBRATION). To shorten the length of turbogenerators, conductor-cooling has proved effective. *See* ELECTRIC ROTATING MACHINERY. *See also* ELECTRIC POWER GENERATION. [L.T.R.]

Bibliography: L. V. Bewley, *Alternating Current Machinery*, 1949; A. E. Knowlton (ed.),

Standard Handbook for Electrical Engineers, 9th ed., 1957.

Genetics

The science that is concerned with biological inheritance—that is, with the causes of the resemblances and differences among related individuals.

Genetics occupies a central position in biology, for essentially the same principles apply to all animals and plants, and an understanding of inheritance is basic for the study of evolution and for the improvement of cultivated plants and domestic animals. It has also been found that genetics has much to contribute to the study of embryology, biochemistry, pathology, anthropology, and other subjects. *See* ANTHROPOLOGY, PHYSICAL; BIOCHEMISTRY; EMBRYOLOGY; PATHOLOGY.

Genetics may also be defined as the science that deals with the nature and behavior of the genes, the fundamental hereditary units (see GENE). From this point of view, evolution is seen as the study of changes in the gene composition of populations, whereas embryology is the study of the effects of the genes on the development of the organism. *See* GENE ACTION; MUTATION (ONTOGENY); POPULATION GENETICS.

The geneticist uses various methods of study, but most all of them rest on the presence of differences between individuals. His techniques do not enable him to analyze those portions of the hereditary make-up of individuals for which he cannot find differences. If differences occur, it is also necessary that they be transmitted to later generations; and in general the full analysis will also require the presence in the material of sexual reproduction or some analogous process that allows the recombination of inherited properties from different individuals. In brief, what the geneticist usually does is to cross diverse individuals and study the descendants (or, in such material as man, analyze existing pedigrees). Such a study must usually be carried through at least two successive generations, and will require enough individuals to establish ratios between the classes present, since conclusions are usually based on these. The mechanism of heredity is such that its analysis requires the use of probability theory, although for much of the subject only very simple statistical concepts are needed. *See* BIOMETRICS; BREEDING (ANIMAL); BREEDING (PLANT); CHROMOSOME; CHROMOSOME ABERRATION; CYTOPLASMIC INHERITANCE; HETEROSIS; HUMAN GENETICS; MENDELISM. [A.H ST]

Gentianales

An order of the plant subclass Dicotyledoneae including 5 families with 524 genera and 5100 species, found mainly in tropical, subtropical, and warm temperate regions. The order is characterized by opposite leaves and the twisted arrangement (estivation) of the petals in the bud. The principal group is the gentian family (Gentianaceae). The dogbane family (Apocynaceae) includes the genus *Rauwolfia*, a source of tranquilizing drugs. The milkweed family (Aesclepiadaceae) usually has milky juice. The milkweeds also have highly specialized, insect-pollinated flowers and pollinia (masses of coherent pollen grains). The olive family (*Oleaceae*) includes olive, ash, lilac, privet, jasmine, and fringe tree; gelsemium and *Strychnos*, source of strychnine and curare, belong to the logania family (Loganiaceae). Allamanda, frangipani, periwinkle, oleander, and *Strophanthus* (source of the drug stropanthin) belong to the family Apocynaceae. *See* ASH; OLIVE; STROPHANTHUS; STRYCHNINE; STRYCHNOS; *see also* DICOTYLEDONEAE; EMBRYOPHYTA; PLANT KINGDOM; TREE. [P.D.S.]

Geochemical prospecting

Use of geochemical data on distribution abundance, mobility, and migration of elements in the search for valuable mineral deposits. An ore body, because of the unusual concentration of a particular element or mineral, is a geochemical anomaly. On exposure to weathering and erosion, the concentrated element or mineral tends to be dispersed from the bedrock source, contaminating adjacent or overlying soils, water, and vegetation, thus developing dispersion patterns. In order to detect geochemical anomalies and dispersion patterns, systematic surveys of the chemical properties of naturally occurring materials are carried out, by a program of careful sampling and analysis. Variations in these chemical properties, considered together with physical and chemical factors controlling the environment and the geologic and geomorphic history of the area, provide significant indications of the presence of concentrations of metals or minerals. *See* ORE AND MINERAL DEPOSITS.

Normal abundances. In order to identify an anomaly, or a significant dispersion pattern derived therefrom, the normal abundance of the element or mineral in the particular environment must be known. In order to establish this normal abundance or background, a large number of samples are collected and analyzed. The values obtained may show a large scatter, but the most frequently recurring values tend to fall within a relatively restricted range about a mode. This modal value is considered to represent the normal abundance or background of the area for the particular material sampled. Samples which contain values of two times background or more may be considered to be anomalous.

Where anomalous values are geometrically grouped so as to form a fairly definite pattern, a geochemical anomaly or a dispersion pattern is present. This anomaly or dispersion pattern may be genetically related to an ore body, and hence, further investigations are warranted. Subeconomic deposits may also give rise to anomalies and dispersion trains. In some cases, certain special features of the environment will favor the development of metal concentrations in surface formations which are completely unrelated to ore-forming processes or mineral deposits. Successful applica-

tion of geochemical techniques depends largely on the ability of the prospector to sort out the anomalies and dispersion patterns which are related to ore deposits from those which are not.

Migration and mobility. In the major geochemical cycle, geologic materials undergo various changes in form and composition as they are subjected to the processes of sedimentation, metamorphism, anatexis, palingenesis, and weathering (*see* LITHOSPHERE, GEOCHEMISTRY OF). During these changes, each element, according to its chemical and physical properties, will tend to be concentrated or dispersed. Elements with similar properties will have similar concentration and dispersion patterns. Physicochemical conditions at depth contrast strongly with those at the surface, and hence elements having similar concentration patterns at depth may have dissimilar patterns at the surface. For example, lead and zinc, which are commonly found closely associated in many ore deposits, are separated in the surficial environment where lead tends to remain close to the source and zinc travels for long distances away from the source.

Geochemical anomalies. The deep-seated (endogenetic) processes of metamorphism and plutonism, or igneous activity, and the surficial (exogenetic) processes of weathering and sedimentation, may both form abnormal metal or mineral concentrations within a host rock. Such concentrations may be termed primary geochemical anomalies if they are formed contemporaneously (syngenetically) with the enclosing rock, and secondary geochemical anomalies if they are formed at some subsequent time (epigenetically). During the process of concentration, a volume of host rock surrounding the area of maximum concentration may be contaminated with small amounts of the ore-forming or related elements. Similarly, during the course of weathering, materials formerly concentrated in the bedrock may be dispersed in the weathered products. Chemical and mineral patterns in bedrock and in surficial materials formed by these dispersion mechanisms are dispersion patterns. *See* ORE DEPOSITS, GEOCHEMISTRY OF.

Primary dispersion patterns. Metallogenetic provinces, dispersion aureoles, and leakage halos yield characteristic primary dispersion patterns.

Metallogenetic provinces. These are portions of the earth's crust which are abnormally rich in certain elements. Such provinces may be indicated by regional dispersion patterns in rocks and minerals developed by processes genetically related to ore formation. Examples are the tin-rich micas of pegmatites in tin provinces and the gold-rich igneous rocks in gold provinces.

Dispersion aureoles. Disseminations of ore-forming elements in wall rocks form dispersion aureoles which are more or less symmetrically disposed about the main zone of [illegible] feet of the ore body and are characteristically developed around replacement deposits in permeable rocks. *See* AUREOLE, CONTACT; METASOMATISM.

Leakage halos. These are dispersion patterns developed along channels and paths followed by mineralizing solutions leading into and away from the zone of concentration. The larger features of the pattern are controlled by the geometry of the ore channels, whereas the detailed distribution of the dispersed metal is controlled by microfracturing and shearing. In such cases, although the detailed geochemical survey data may appear to yield erratic results, statistical analysis may serve to reveal the larger pattern. Leakage halos have been identified in fracture zones up to distances of 500 ft from the ore bodies, and in shear zones up to distances of several thousand feet. Gaseous dispersion patterns, such as those related to concentrations of natural gas and oil, are true leakage halos.

Secondary dispersion patterns. Secondary dispersion patterns are developed in inorganic and organic surficial materials by the release and redistribution of elements and minerals upon weathering of exposed mineral deposits. The extent and nature of the dispersion pattern are dependent in large part on the mobilities of the elements; these in turn are dependent on the physical, chemical, and biological characteristics of the environment. The products of weathering move down-slope under the influence of gravity, and their rate of movement is directly proportional to the amount of water present.

The most mobile substances are those that form a gas phase on weathering. Slightly less mobile are those that form ionic or colloidal aqueous solutions and suspensions. The least mobile are those that are relatively insoluble and have a high specific gravity. The movements of this last group are controlled almost entirely by the physical factors of the environment. Movements of soluble substances are controlled largely by chemical factors, such as variations in pH and Eh (oxidation-reduction potential) of waters and soils, and by biologic factors, such as the presence or absence of vegetation and bacteria. Variations in the biological and chemical environment may result in removal of a mobile substance from solution by precipitation, adsorption, or organic assimilation, thus seriously restricting the development of a large dispersion pattern.

The empirically determined mobility of some of the ore-forming elements, in order of increasing mobility, is as follows: Fe, As, Pb, Au, Cu, Co, Zn, and Ag. This order of mobility is valid only for the relatively acid environment in the vicinity of sulfide-rich deposits. An alkaline environment, or the presence of chlorides or carbonates, seriously restricts the mobility of such ions as copper and silver.

Secondary dispersion patterns may be divided into two groups, those involving little or no lat-

eral migration of a dispersion element or mineral, and those in which lateral migration is of major importance.

The first group includes residual and vestigial dispersion halos and saline dispersion halos. Residual and vestigial dispersion halos are those in which abnormal amounts of elements and minerals remain in the residual soils derived from a bedrock source. The source may still be present below the halo, or it may have been at some higher level. Saline dispersion halos are patterns developed where concentrations of mobile elements have moved upwards into transported soils above a bedrock source. Dispersion patterns of this group are best developed in flat terrains.

The second group includes dispersion fans and dispersion trains. Dispersion fans develop where the products of weathering have an asymmetrical distribution diverging from the source, such as in the case of movement by glaciers or by mass wasting. Dispersion trains are linear dispersion patterns leading away from the source, and are normally controlled by the surface drainage system.

In detail, any of the above patterns may be homogeneous or heterogeneous. Patterns developed by the movement of materials in solution are relatively uniform and sampling presents few difficulties. Patterns developed by mechanical processes of transportation, such as glaciation, tend to be erratic, and careful sampling combined with a statistical analysis of the results may be required to make the pattern clearly evident.

An interpretation of the patterns produced by secondary dispersion processes must take into consideration all environmental conditions in the area. Some of these controlling factors are the bedrock geology, the geomorphic characteristics and history of the area, ground-water and surface-water movement, type and distribution of vegetation, the physicochemical characteristics of waters, soils, and alluvium, and possible sources of artificial contamination.

Geochemical survey techniques. The prime requirement for geochemical surveys is an adequate sampling program. Depending on circumstances, the material sampled may be a selected horizon in transported or residual soil, stream sediment, surface or ground water, or vegetation growing in or on these materials. Where possible, sampling should be confined to one type of material only. When diverse materials must be sampled because of the heterogeneous nature of the terrain or ecology, correction factors must be applied to the results before the geochemical patterns can be interpreted. The spacing and distribution of the sampling points are controlled by the estimated size of the metalliferous source, the mobility of the dispersing metals, and the type of dispersion pattern expected.

Geochemical surveys have been classified, according to the type of material sampled, as soil surveys, sediment surveys, water surveys, and vegetation surveys.

Soil surveys. In these surveys, a suitable soil horizon is selected and sampled throughout the area. Such horizons develop under relatively constant physicochemical conditions in any given climate, and hence, have fairly uniform properties. Optimum sensitivity can be achieved by sampling a horizon in which the dispersed elements tend to accumulate. For chemically dispersed elements, the B horizon of the podzolic soils is most satisfactory. In areas of organic soils such as peat, muskeg, and others, the basal humified portion of the organic layer is usually taken. For mechanically dispersed resistate (stable) minerals, the basal part of the C horizon should be sampled. Sampling positions are normally located on a rectangular grid pattern. *See* SOIL; SOIL (GREAT SOIL GROUPS).

The samples are dried and sieved, and a fraction of suitable size is selected for analysis. For chemically dispersed elements, the fine fraction (−80 mesh) is suitable, whereas for resistate minerals a coarser fraction is more satisfactory.

Sediment surveys. The purpose of these surveys is to detect the migration path of dispersed elements and minerals along the surface drainage channels in an area. Samples are collected from the fresh sediment in the bottom of the channels and also from old sediment on the terraces and flood plains. For chemically dispersed material, the fine fraction of the sample is used for analysis; for mechanically dispersed material, a coarser fraction is selected. Sampling points are located at intervals along the length of the drainage system. Sediment surveys are best carried out at low-water periods when channel bottoms and flood plains are readily accessible.

Water surveys. Either ground waters or surface waters may be tested, depending on local conditions. Water surveys are suitable only for detecting soluble materials, and hence are restricted to the search for chemically dispersed elements or salts. Seasonal and diurnal fluctuations of the metal content of waters have been observed and appear to be directly related to variations in the amount of local precipitation, which is accompanied by an elevation or depression of the water table. The metal content of waters is of the order of a few parts per 10^9 parts and hence very sensitive analytical techniques must be used or large volumes of samples must be tested. To avoid problems introduced by transportation and contamination of water samples, field analyses are made when possible.

Surface waters are sampled at regular intervals along the drainage net; ground waters can only be sampled on an irregular pattern dependent on the distribution of wells, springs, and seeps.

Vegetation surveys. These may be either biogeochemical surveys or geobotanical surveys. In biogeochemical surveys, the trace-element content of selected plant material is determined by

cal analysis. Samples are collected from individuals of the same species and from vegetation of the same type and growth. Satisfactory results have been obtained by using second- and third-year growth from twigs and branches. Care must be taken to determine that variations in metal content of the plant are reflections of changes in the exchangeable metal content of the soils and do not merely reflect changes in growth factors not related to metallization. These growth factors are related to, and controlled by, such variables as drainage conditions, microclimate, and soil pH. *See* ECOLOGY.

In geobotanical surveys, the distribution of indicator plant species, or plant symptoms diagnostic of high metal soils, are mapped from field observations. When suitable indicator plants are present, this method is rapid and inexpensive, but unfortunately such plants do not appear to be consistent from one area to another. Geobotanical techniques require careful preliminary orientation surveys carried out by trained personnel. *See* COMMUNITY; ECOLOGY, PHYSIOLOGICAL.

Analytical techniques. In most geochemical surveys, it is desirable to differentiate between total metal in the sample and loosely bonded or exchangeable metal. Therefore, a preliminary extraction of appropriate intensity is carried out, and the metal content of the extract is then determined by chemical or instrumental techniques. A wide variety of analytical methods has been used, including colorimetric and chromatographic analyses, polarography, and emission spectrography. When it is desirable to determine the total metal content of the sample, optical spectrography and x-ray fluorescence spectrometry are most widely used. Optical spectrographic techniques are used for multicomponent analyses except for a few elements with poor emission characteristics. Most field tests are based on colorimetric methods.

An analytical technique should be sufficiently sensitive, and yet be inexpensive, and unaffected by changes in bulk composition of the sample material. Portability is of prime importance in the case of field tests.

When mineral dispersion patterns are sought, the coarser fraction of the sample is separated by panning or sieving, and examined by optical or x-ray diffraction techniques.

Field program. A full-scale geochemical prospecting program would include these stages:

1. *Preliminary* evaluation of areas selected on the basis of available geologic data, by sampling and testing of intrusive and sedimentary rocks and minerals theoretically related to the type of metallization. In this way, a metallogenetic province can be identified.

2. Primary reconnaissance and orientation surveys, based on sampling of major drainage basins, using sediment surveys.

3. Secondary reconnaissance surveys based on detailed testing of drainage basins containing anomalous values. Poorly drained areas should be tested by widely spaced sampling of soil and ground waters.

4. Follow-up surveys along dispersion trains to determine cut-off points and extent of dispersion patterns. These surveys are normally a combination of sediment, water, and soil testing. Priority for follow-up surveys would be based on the presence of favorable rocks and structures, favorable geophysical indications, and intensity of the geochemical anomaly.

5. Detailed surveys carried out in the vicinity of the metalliferous source by sampling of soil at closely spaced intervals. The results of these surveys would be interpreted in terms of bedrock mineralization, only after taking into consideration factors which might influence the mobility of the metal. Geological and geophysical data are useful in the interpretation. *See* MINERAL RESOURCE AREAS; *see also* GEOPHYSICAL EXPLORATION; PROSPECTING.

[J.E.R.]

Bibliography: I. I. Ginsburg, *Principles of Geochemical Prospecting*, 1959; J. Green, Geochemical table of the elements for 1959, *Bull. Geol. Soc. Am.*, 70:1127–1183, 1959; H. E. Hawkes, *Principles of Geochemical Prospecting*, USGS Bull. 1000-F, 1957; K. Rankama and Th. G. Sahama, *Geochemistry*, 1950.

Geochemistry

The study of the chemical composition of the various phases of the earth and the physical and chemical processes which have produced the observed distribution of the elements and nuclides in these phases. Classical geochemistry was directed largely to defining the abundance of the elements in rocks and minerals. Modern techniques, the most important employing isotopic dilution and neutron activation, have greatly extended the sensitivity and accuracy of geochemical analyses. Phase equilibria of simple petrologic systems have been studied in the laboratory under high pressure and temperature. These studies have advanced man's knowledge of the nature of chemical reactions in the crust. Similar work with sulfide systems is providing new data about the origin and emplacement of ore deposits. Studies of the chemistry of organic constituents in marine sedimentary rocks have been made in an attempt to understand the formation of oil pools. Trace-element distribution in surface waters, soils, and plants is now used to detect ore deposits. *See* ATMOSPHERE, GEOCHEMISTRY OF; BIOSPHERE, GEOCHEMISTRY OF; ELEMENTS (COSMIC ABUNDANCE); ELEMENTS (GEOCHEMICAL DISTRIBUTION); GEOCHEMICAL PROSPECTING; HIGH PRESSURE PHENOMENA; HYDROSPHERE, GEOCHEMISTRY OF; LITHOSPHERE, GEOCHEMISTRY OF; ORE DEPOSITS, GEOCHEMISTRY OF; PETROLEUM (ORIGIN); SILICATE PHASE EQUILIBRIA; SULFIDE PHASE EQUILIBRIA; TERRESTRIAL NUCLEAR REACTIONS. *See also* ACTIVATION ANALYSIS; ISOTOPE DILUTION TECHNIQUES; PETROLOGY.

Modern geochemical research places considerable stress on the quantitative determination of temperature, pressure, and time in geologic systems. Methods based on isotopic fractionation and radioactivity have proved extremely useful in this regard. Most of these developments have occurred since 1950. In the last decade, the impact of isotopic geochemistry in earth science has been as important as the discovery of stratigraphic correlation with fossils or the determination of subsurface structure with the seismograph and gravity meter. In November, 1955, an international organization, the Geochemical Society, was formed and *Geochimica et Cosmochimica Acta* was adopted as the official journal. See GEOCHRONOMETRY; LEAD ISOTOPES, GEOCHEMISTRY OF. See also COSMOCHEMISTRY; PALEOBIOLOGIC THERMOMETRY; LEAD ISOTOPES, GEOCHEMISTRY OF. See also COSMOCHEMISTRY. [J.L.K.]

Bibliography: P. Abelson (ed.), *Researches in Geochemistry*, 1959; B. Mason, *Principles of Geochemistry*, 2d ed., 1958; A. Muir (ed.), *Geochemistry*, 1954; K. Rankama and T. G. Sahama, *Geochemistry*, 1950.

Geochronometry

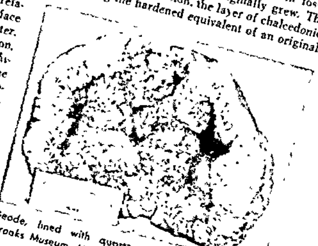
The study of the absolute age of the rocks of the earth using methods based on the radioactive decay of such isotopes as U^{238}, U^{235}, Th^{232}, Rb^{87}, K^{40}, and C^{14}. Measurements using these decay schemes permit definition of the age of the earth and meteorites (4.5×10^9 years); the oldest available rocks (3×10^9 years); significant events in the geologic time scale; the rate of evolution of organisms; the time and duration of major events in crustal evolution; the advance and retreat of continental glaciers; and the evolution of human culture. These isotopic chronometers have superseded many of the earlier qualitative methods for estimating geologic time and their application introduces new quantitative data into earth science. See EARTH (AGE OF); GEOLOGICAL TIME SCALE; LEAD ISOTOPES, GEOCHEMISTRY OF; RADIOCARBON DATING; ROCK (AGE DETERMINATION). [J.L.K.]

Bibliography: J. L. Kulp, *Quantitative Geochronometry*, 1960.

Geode

A roughly spheroidal hollow body, lined on the inside with inward-projecting small crystals. Geodes are found most frequently in limestone beds but may occur in some shales. Typically, a geode consists of a thin outer shell of dense chalcedonic silica and an inner shell of quartz crystals, sometimes beautifully terminated, pointing toward the hollow interior (see CHALCEDONY). Many geodes are filled with water; others, having been exposed for some time at the surface, are dry. Calcite or dolomite crystals line the interior of some geodes, and a host of other minerals are less commonly found. In some geodes there is an alternation of layers of silica and calcite, but almost all geodes show some banding suggestive of rhythmic precipitation.

The origin of geodes lies in the presence of an original cavity, in many cases voids within fossil shells, from which the geode originally grew. The geode grows by expansion, the layer of chalcedonic silica being the hardened equivalent of an original



Geode, lined with quartz crystals, Keokuk, Iowa. (Brooks Museum, University of Virginia)

silica gel. The expansion is due to osmotic pressure from original sea water trapped inside the silica gel shell and fresh water on the outside of the gel. The projecting quartz crystals are precipitated from later ground waters infiltrating the already hardened, hollow spheroid. See SEDIMENTARY ROCKS. [B.S.]

Geodesy

An earth science that aims to determine the size and shape of the earth, to study the earth's gravity field on and above the earth's surface and, in conjunction with other earth sciences, to study the structure of the earth's crust and of the immediate substrata. Its practical purpose is to perform the measurements and computations needed to determine the coordinates of the control points that are necessary to sound mapping and charting.

Dimensions of the earth. It is useful to recognize three types of earth surface: the physical surface with continents and oceans, mountains, valleys, and low lands, of interest to the geographers; the mathematical surface, or earth ellipsoid of revolution, along which the geodetic computations are carried out; and the geoid, the level, or equipotential, surface of the earth's gravity. The geoid, as the level surface, is perpendicular to the plumb line. In the oceans, the geoid coincides with the average sea level.

The earth ellipsoid is flattened at the poles because of the centrifugal force caused by the earth's rotation around its axis. The ellipsoid is completely defined, when its two parameters are known: the equatorial radius a and the flattening α of the meridians. Flattening α is by definition $\alpha = (a - b)/a$, where b is the polar radius of the e

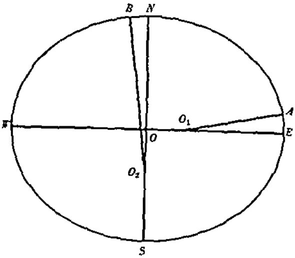


Fig 1. Diagrammatic representation of the earth ellipsoid. ON is the polar radius, b; OE, equatorial radius, a; O_2N and O_1E are radii of curvature of the meridian at the poles and at the equator; thus $\alpha = (a - b)/a$. Note that $O_1A < b < a < O_2N$.

(Fig. 1) From a and α, the radius b can be computed.

The table gives the dimensions of the earth ellipsoid, mostly used in geodesy, with the authors and computation year. Different scientists have obtained different values for a and α (*see* TERRESTRIAL GRAVITATION).

Dimensions of the earth

Author	Year	a, meters	$1/\alpha$
Bouguer, Maupertuis	1738	6,397,300	216 8
Clarke, A B	1866	8,206	295 0
Clarke, A R.	1880	8,249	293 5
Hayford (International Ellipsoid)	1910	8,388	297 0
Heiskanen, W A	1926	8,397	297.0
Krassowski, F N	1938	8,245	298 3
Jeffreys, H.	1948	8,099	297.1
U S Army Map Service (Hough)	1956	8,260	297 0

The polar radius b is about 21.5 km shorter than the equatorial radius a. On a globe of radius 1 m, b would be 3 mm short...

...standard, but during recent decades, most countries adopted the ellipsoid computed by the American ... 1924 at the ... al Union of ..., as the International Ellipsoid. The Soviet Union uses the Krassowski ellipsoid of 1938.

Last computations obtained by the U.S. Army Map Service on the basis of a huge mass of material indicate that it may be desirable to reduce the a value of the international ellipsoid by about 130 m and the α value by about 0.3%, and also to use the f value of 1/298.0 instead of 1/297.0. Hints in the same direction also appear in the results obtained earlier in Russia and in the late 1950s ...

for two problems: the geodetic problem, the measurement of an arc l, or better, several arcs, along a great circle of the earth such as along a meridian; and the astronomic problem, to determine the central angle v corresponding to the measured arc.

Figure 2 illustrates schematically the arc measurement of Eratosthenes and explains that the radius r of the earth—assumed to be a sphere—is $r = l/\omega$, where ω is given in radians. The longer l is the larger ...

the equatorial radius a and the flattening α.

This method, called the arc-measuring method, was used between 276 and 194 B.C. by Eratosthenes, the first to measure the size of the earth. Instead of the camels and Syene's well that Eratosthenes used in his classic determination of the size of the earth, however, today's techniques include triangulation, Shoran, Hiran, and other electronic methods; and different celestial methods, including artificial satellites and the finest astronomic telescopes and transits, to obtain as accurate values as possible for a and α (*see* SURVEYING). This article considers only triangulation, the basic method of geodesy.

Triangulation. The principle illustrated by Fig. 3 is based on the simple fact that a triangle is completely defined when three quantities are known, of

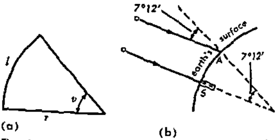


Fig. 2. Arc-measuring method. (a) Principle of the arc-measuring method. (b) Diagram of Eratosthenes' method of 2200 years ago using camel caravan to measure distance SA (Syene-Alexandria) and the well at Syene to aid in measuring the central angle. (*From W. A. Heiskanen and F. A. Vening Meinesz, The Earth and Its Gravity Field, McGraw-Hill, 1958*)

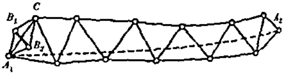


Fig. 3. Principle of triangulation. When the distance B_1B_2 (base line) and the angles of all triangles are measured, it is possible to compute all sides of the triangles and the distance A_1A_2 as well. (*From W. A. Heiskanen and F. A. Vening Meinesz, The Earth and Its Gravity Field, McGraw-Hill, 1958*)

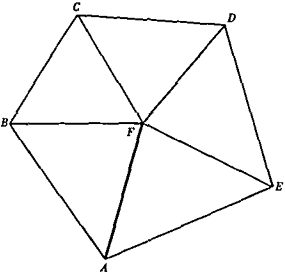


Fig 4. Adjustment of the central system. When the angles of the pentagon $ABCDEF$ are measured, it is possible to compute the side AF in applying sine theorem for all five triangles. The end value of AF should equal the starting value.

which one must be a side or other linear quantity. The other two can be angles. In a triangulation chain or net, in principle one need measure only one side, A_1C, enlarged from a carefully measured base line B_1B_2. From that can be computed the starting side A_1C, and from that, all other sides of this triangle and other triangles until the end point A In fact, more base lines are commonly measured for control, in general, at intervals of about 250 km.

It is not necessary to measure more than two angles, α and β, in any triangle. The third, γ, can be obtained from the formula

$$\alpha + \beta + \gamma = 180° + \epsilon$$

where ϵ is a small quantity, called the spherical excess, brought about by the curvature of the earth. All three angles of any triangle are measured as a check on every independent triangle. This is the basis of a control, or a condition, equation

$$\alpha + \beta + \gamma - (180° + \epsilon) = w$$

where w is the closure error of the triangulation. Such equations are called triangle equations.

In every central system, such as illustrated in Fig. 4, a condition equation may be established and called a side equation. In computing, for instance, the sides of all five triangles from the side AF and thence around the central point, AF must check with its original length. The difference between the original and end values of AF is the closure of the side equation.

Between two base lines, a closure, called base line equation, may also be established. If there are n base lines in a triangulation, there will be $(n-1)$ independent side equations.

Every triangulation loop from a point A back to it yields three closures, or so-called coordinate equations: one in latitude, one in longitude, and one in azimuth.

On the basis of the method of least squares, the observed angles are corrected in a geodetic adjustment so that all closures disappear, and at the same time, the square sum Σv^2 of the corrections v to the angles or directions is a minimum. Then, it becomes evident that the corrections are most plausible and that all inner discrepancies disappear between the observations. Of course, observation errors are not eliminated by such a check; it is possible only to lessen inner discrepancy. The more reliable end results that can be obtained, the better the observations. Even the best adjustment will not make good ones from poor observations.

In triangulation, it is necessary to make angle observations, base line measurements, and azimuth observations. To change the usual triangulation in arc measurement, it is still necessary to measure the central angles corresponding to different arcs, or to measure, astronomically at least, latitude φ or longitude λ or azimuth A at as many triangulation points as possible. The astronomic latitude observations give, together with the geodetically computed coordinate φ', the vertical deflection component ξ in meridian; the astronomic longitude λ' or azimuth A' observations give the vertical deflection components η in the east-west directions. If both longitude and azimuth are observed astronomically, two values are obtained for η and a Laplace equation relating them. These astrogeodetic points are called the Laplace points. Relatively few triangulation points are Laplace points, except in Finland, where almost all triangulation points of first order are Laplace points.

The important vertical deflection components read

$$\begin{cases} \xi = (\varphi' - \varphi) \\ \eta = (\lambda' - \lambda) \cos \varphi \\ \eta = (A' - A) \cot \varphi \end{cases}$$

The two last equations give the Laplace equation between the vertical deflections in longitude and azimuth:

$$A' - A = (\lambda' - \lambda) \sin \varphi$$

The vertical deflection equations play a central role in the astrogeodetic method for determination of the dimensions of the earth.

Curvature of the earth's surface. The curvature of a surface at a certain point is the inverse value $1/r$ of the radius r of the curvature. For a sphere's surface, both r and $1/r$ are everywhere the same, but for the earth ellipsoid they vary not only with the latitude but also with the azimuth. Only along the same latitude parallel is r the same, and consequently the curvature remains constant. In geodesy, the meridian radius of the curvature R and the east-west radius of curvature N are fundamental figures for the computation of the coordinates. They can be obtained from the formulas

$$R = \frac{a(1-e^2)}{W^3} = \frac{c}{V^3} \quad \text{and} \quad N = \frac{a}{W} = \frac{c}{V}$$

in which $c = a^2/b$ is radius of curvature at the pole, $e = \sqrt{a^2 - b^2}/a$, $e' = \sqrt{a^2 - b^2}/b$ are the eccentricities of the meridian, and

$$W^2 = (1 - e^2 \sin^2 \varphi) \qquad V^2 = (1 + e'^2 \cos^2 \varphi)$$

The equations

$$(1 - e^2)(1 + e'^2) = 1 \quad \text{and} \quad e^2 = \alpha(2 - \alpha)$$

give the relationship between e^2, e'^2, and the flattening α of the meridian. Except at the poles, where R and N are both equal, N is larger than R, the ratio being $(1 - e^2 \sin^2 \varphi)/(1 - e^2)$. The mean radius of curvature

$$r = \sqrt{RN} = \frac{a\sqrt{1 - e^2}}{W^2} = \frac{c}{V^2}$$

The radius of curvature r_a at the azimuth α can be obtained from the formula

$$1/r_a = (\cos^2 \alpha)/R^2 + (\sin^2 \alpha)/N^2$$

Because the geoid is an irregular surface, its radius of curvature changes irregularly from point to point. Its value is relatively small in these considerations and therefore is neglected.

Geographic location. The location of a certain point on the earth's ellipsoid is given by the coordinates latitude φ and longitude λ. Latitude is computed from the Equator positive to $+90°$ at the North Pole and negative to $-90°$ at the South Pole. Longitude is in general computed from the meridian through Greenwich, positive eastward to 180° and negative westward to the same longitude of 180°. In many cases longitude is denoted by λ_E (eastward) and λ_W (westward). Longitude can also be given in time units from 0^h to 24^h; 1^h = 15°, or 15° rotation per hour.

It is customary to use only one kind of longitude, λ, but there are in common use three different types of latitudes: (1) the geographic or geodetic latitude φ, which is equal to the polar height (Fig. 5) and can be observed astronomically or geodetically; (2) geocentric latitude γ, the angle between the Equator and the geocentric radius OP, and (3) the reduced latitude ψ, important in geodetic applications. Complex relations between γ, ψ, and φ are expressed in several technical formulas:

$$\tan \gamma = \sqrt{1 - e^2} \tan \psi = (1 - e^2) \tan \varphi$$

Thus, $\varphi > \psi > \gamma$, except at the poles, where all are 90° and at the Equator where all are 0°. Maximum difference $\varphi - \gamma = 11.6'$ is at the latitude $\varphi = 45°$.

Coordinate systems. Most used in geodesy are the geographic coordinates: latitude φ and longitude λ, and the rectangular coordinates x and y. The value of x is computed from the Equator positive northward to the North Pole and negative southward to the South Pole; y is computed from the Greenwich meridian, positive eastward, negative westward. In England and America, x and y are oppositely used. For English geodesy, x is the east-west coordinate, y the north-south coordinate.

Rectangular coordinates differ depending on what projection is used. In geodesy, conformal projections are used almost exclusively. The main characteristic of all conformal projections is the fact that the shapes of the different small parts of the map remain relatively undistorted in the projection. In other words, the scale factors m_1 in the meridian directions and m_2 in the east-west direction are everywhere equal. Both of them, however, change from point to point. The most-used geodetic map projections are the Mercator projection, the Transverse Mercator projection, and the Lambert Conical conformal projection. In sea charts, the Mercator projection or a modification of it is used almost exclusively. For maps of the continents, the Transverse Mercator projection is most used. In the United States, the Transverse Mercator projection and the Conical Lambert projection are almost equally used.

Earth's moments of inertia. The moment of inertia C of the earth around the rotation axis is by definition

$$\int_m (x^2 + y^2)\, dm$$

and the moment of inertia A around the arbitrary axis in the equatorial plane is

$$\int_m (r_a^2 + z^2)\, dm$$

in which x and y are the equatorial rectangular coordinates, z is the distance of the mass element dm from the Equator, and r_a its distance from the X or Y axis of the Equator. The integration has to be extended over the mass M of the whole earth. If the earth is considered a triaxial ellipsoid, the moments of inertia A and B around the Y axis and X axis are different. The moments of inertia of the earth have certain technical relationships to geodesy and geophysics.

Dynamical ellipticity. Since the earth is flattened at the poles, the moment of inertia C must be larger than the moment of inertia A. The difference of

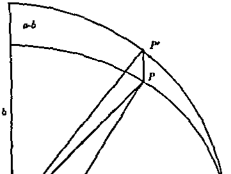


Fig. 5. Illustration of the three types of latitudes: the geographic latitude φ, the reduced latitude ψ, and the geocentric latitude γ.

$(C-A)$ must depend on the flattening or the ellipticity ϵ of the earth. The ratio $\epsilon = (C-A)/C$ is called the *dynamical ellipticity* of the earth. It depends on the shape of the earth and on the mass distribution in the earth. The perturbations of the moon's orbit brought about by the flattened earth give the quantity $(C-A)/Ma^2$, and the precession in combination of the mass of the moon gives the quantity $(C-A)/C$, or dynamical ellipticity ϵ. By this method H. Jeffreys obtains the value $\epsilon = 1/297.10$ or close to the flattening $\alpha = 1/297.0$ obtained gravimetrically and astrogeodetically.

The celestial method of Jeffreys applied to the perturbation of the artificial satellites has given the value $\epsilon = 1/298.15$.

The ratio C/Ma^2 is, according to Jeffreys, 0.334 ± 0.002, and according to E. C. Bullard, 0.335. If the earth were homogeneous, this ratio would be 0.4.

Variation of latitude. Because the earth is flattened and therefore the difference $(C-A)$ of the principal moments of inertia is not zero, the instantaneous axis of rotation moves in a circular cone about the axis of figure of the earth. Its period of revolution is $A/(C-A)$ rotation periods or about 305 days, and is called Euler's period because he first predicted the existence of this phenomenon. Because the instantaneous pole rotates around its normal position, the latitudes of the earth, which refer to the instantaneous pole, change periodically. This movement is called the variation of latitude, found first by F. Küstner in 1888. By careful studies, S. C. Chandler in 1892 was able to show that the observed variations, with amplitudes of the order 0.″1, can be explained by two terms, one of period 14 months, the other of period 12 months. The annual portion is obviously caused by meteorological effects. S. Newcomb realized that

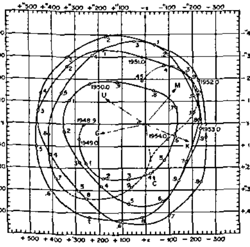


Fig 6. Observed movement of the North Pole. The irregular movement of pole is eccentric with the normal position of the pole. Its amplitude is rather small, never larger than 0.″47. (*After G. Cecchini in B. F. Howell, Jr., Introduction to Geophysics, McGraw-Hill, 1959*)

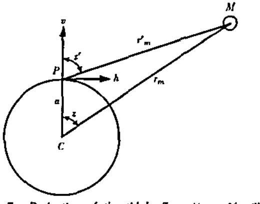


Fig. 7. Derivation of the tidal effect. Moon M will cause tidal effect with the horizontal component h and vertical component v. The sun causes a similar effect. *The period of the main tidal waves is about half a day. (From W. A. Heiskanen and F. A. Vening Meinesz, The Earth and Its Gravity Field, McGraw-Hill, 1958)*

the 14-month period is only a modified Eulerian motion. The period 305 days of Euler's theory suits for the absolutely rigid body, but for the actual elastic and plastic earth, the period increases by about 2 months; that is, the more precise result is Chandler's period of 429 days.

Figure 6 represents the motion of the instantaneous and shows how irregular this motion is. The amplitude varies, but is always under the limits $\pm 0.''45 = +13$ m. The pole variation has considerable effect upon azimuth observations at high latitudes because this effect increases with the tangent of the latitude.

Tides. The attraction of the mass of moon or sun is larger at the surface point of the earth closest to the moon or sun, and smaller at the point farthest from the moon or sun than at the earth's center. Consequently, the water masses of the oceans must be periodically higher and lower. This periodical motion of the ocean surface, brought about by the variable attraction of these celestial bodies, is called the ocean tides, and the perturbation force of moon and sun is called the tidal force. It can be obtained at any point P, Fig. 7, of the earth's surface by subtracting the attraction of moon and sun from the attraction of the same bodies at the earth's center. It has the horizontal and vertical components h and v which read

$$h = \frac{3}{2} g \frac{m}{M} \frac{a}{r^3} R^2 \sin 2z$$

$$v = 3g \frac{m}{M} \frac{a}{r^3} R^2 \left(\cos^2 z - \frac{1}{3}\right)$$

in which M is the mass of the earth, g the gravity, a the earth radius in P, R the average radius of the earth, and z the geocentric zenith distance of M. The quantities m (mass), r (distance), and z refer either to moon, M, or to sun.

The tidal force has effect also upon gravity fact, this correction, Δg to the gravity, $\Delta g =$ is 0.11 mgal for the moon, 0.05 mgal for

and 0.16 mgal for the combined effect during the time of the maximum tide. This correction must be considered in the more accurate gravity observations.

The earth tides. If the earth's interior were completely rigid, the earth's crust would not move at all with the periodical tidal force of moon and sun. Since, however, the earth is partly plastic and elastic (0.28 of the theoretical value of a completely plastic or elastic earth), it participates in the tidal motion, although in much smaller amount than the ocean waters. Tidal effect can now best be measured by gravimeters. *See* EARTH TIDES; TIDE.

[W.A.H.]

Bibliography: G. Bomford, *Geodesy*, 1952; W. A. Heiskanen and F. A. Vening Meinesz, *The Earth and Its Gravity Field*, 1958; B. F. Howell, Jr., *Introduction to Geophysics*, 1959.

Geography, mathematical

The branch of physical geography that deals with

the face of the earth.

The earth is a nearly spherical body, 24,901.92 miles in circumference, rotating on an axis, and revolving in an orbit around the sun. The equatorial diameter is 7926.68 miles and the polar diameter is 7899.98 miles. Its shape may be better classed as an oblate spheroid. Departure from a true sphere, in addition to its shorter polar diameter, is apparent in the mountains and continental masses above the oceans and the deep ocean basins. *See* GEODESY.

The exact nature of all earth materials is not known, especially that of the earth's core. Although the earth appears rigid, and probably is about as rigid as steel, there is a certain element of plasticity in its nature. This is shown by the slight flattening at the poles, caused by centrifugal force, and by the apparent inability of the interior to support much larger masses above the surface than exist at the present time. It is believed that these masses, or continents, are made of lighter materials, such as granite, and that the basins under the oceans are of heavier materials similar to diabase or basalt. The state of balance between the two is constantly being adjusted by such forces as earth movements, volcanism, and the removal of materials from the continents by streams. Enveloped in air (atmosphere), approximately 71% of the earth's outer rock-shell (lithosphere) is covered by large bodies of water (hydrosphere). Only about 29%, therefore, of area is dry land exposed above water. *See* GEOLOGY.

Rotation and revolution. These are the two major motions of the earth. The earth rotates on an axis, the ends of which are the poles, making one

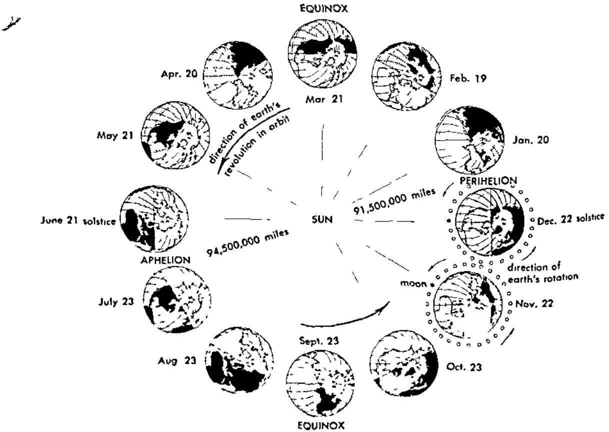


Fig. 1. A diagrammatic plan of the earth's revolution around the sun, showing selected lunar relations and the relation of axial inclination and parallelism to the seasonal changes. (*After A. K. Lobeck, The Earth in Space, C. S. Hammond, 1929*)

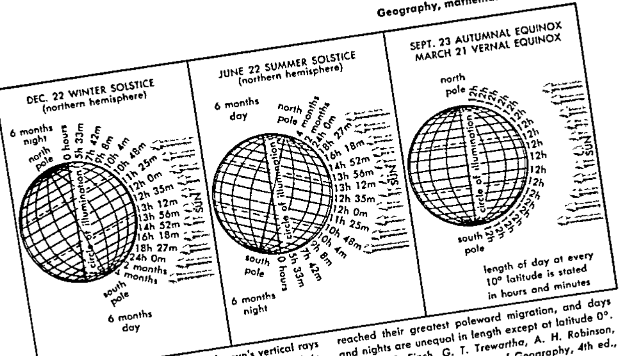


Fig. 2. On the equinoxes, when the sun's vertical rays are at the Equator, days and nights are equal in length over the entire earth. At this time insolation decreases regularly from the Equator to poles. At the times of the solstices, the sun's vertical rays have reached their greatest poleward migration, and days and nights are unequal in length except at latitude 0°. (After V. C. Finch, G. T. Trewartha, A. H. Robinson, and E. H Hammond, Elements of Geography, 4th ed., McGraw-Hill, 1957)

complete rotation in 24 hours, with a speed at the Equator of about 1000 mph. The rotation is from west to east (in earth-bound terms), giving us night and day, and the impression that heavenly bodies rise in the east and set in the west.

The rotating earth revolves in an orbit around the sun, making one complete revolution in 365 days, 5 hours, 48 min, 45.15 sec, or about 365¼ days, thus establishing the length of a year. Every fourth year (leap year) an extra day is added in February to allow for the accumulated quarter days. The earth's path around the sun is not circular, but elliptical, and the plane of the course it follows is referred to as the plane of the ecliptic.

Because of the ellipticity of its orbit, the earth, throughout the year, is at different distances from the sun. Its closest approach is on January 3, when the earth is approximately 91,500,000 miles distant (perihelion). It is farthest from the sun on July 4, about 94,500,000 miles distant (aphelion). The speed of the earth in its orbit is about 66,000 mph, but it varies in the different parts of the ecliptic path, being greatest at perihelion and slowest at aphelion This change in speed produces variations in the length of days through the year.

Axial inclination and parallelism. The positions and changes of the earth's axis are most important factors in earth-sun relationships. As the earth moves through its orbit, the axis of rotation is constantly inclined at about 66½° (66°55) from the plane of the ecliptic, or 23½° (23°45) from a line perpendicular to the ecliptic. In addition to maintaining its inclination with the ecliptic, the polar axis of rotation always points to the same spot in the heavens throughout its yearly revolution around the sun. Thus, there is always parallelism of the axis at all points in the orbit. Precessional motion is not considered here. *See* PRECESSION OF EQUINOXES.

Because of the inclination and parallelism the north polar part of the axis will, on one day of the year (December 22), be tilted 23½° away from the sun; 6 months later, on June 21, it will be tilted 23½° toward the sun. On the two days midway between these extremes, March 21 and September 23, the North Pole will not tilt toward or away from the sun, but the axis will make a 90° angle with a line drawn to the sun.

Thus it can be seen that the vertical rays of the sun (at noon in all cases cited) will continually change position and move through a latitudinal span of 47°, from the Tropic of Cancer 23½° north to the Tropic of Capricorn 23½° south. This progression of the vertical rays of the sun is the basis for the seasons on the face of the earth.

Seasons. From June 21, when the sun's rays are vertical at the Tropic of Cancer (summer solstice), summer continues in the Northern Hemisphere until September 23, when the sun rays strike directly over the equator (autumnal equinox). During this period (and during the spring) the sun at the North Pole will be at or above the horizon at all times. All places north of the Equator will have longer days than nights, and the Northern Hemisphere will receive its maximum solar insolation. Similar summer conditions will be found during December, January, and February at the South Pole and throughout the Southern Hemisphere (*see* NORTH POLE; SOUTH POLE). Fall begins with the sun overhead at the Equator and continues until December 22 (winter solstice), when the sun has its southern zenithal position at the Tropi

of Capricorn, 23½° south. Winter finds the overhead sun rays moving back toward the equator, which is attained on March 21 (vernal equinox). During the fall and winter, days in the Northern Hemisphere will be shorter than the nights and the sun will drop to its lowest angle above the horizon. Spring is that period when the sun is again overhead north of the Equator, from March 21 to June 21. *See* EQUINOX; SOLSTICE.

Earth-time relations. The system of time used by man is also a result of earth motions. It takes the earth 24 hours to make one complete rotation of 360°, or it rotates through 15° of longitude in 1 hour. When it is noon at a given meridian it will be 1 hour past noon 15° to the east and 1 hour before noon 15° to the west. *See* EARTH (ORBITAL MOTION); INTERNATIONAL DATE LINE; TIME.

Lunar-terrestrial relations. Although less important to man than earth-sun relationships, the moon plays a significant role as the major force in the production and timing of tides. The moon (diameter about 2160 miles) is considerably smaller than the earth but its movements are quite similar; its orbital revolution around the earth and its axial rotation are in the same direction. The distance from earth to moon varies from about 222,000 miles at its closest approach (perigee), to about 253,000 miles at its greatest distance (apogee). The moon revolves around the earth once in about 27⅓ days (27 days, 7 hours, 43 min, 11.5 sec); this is called the sidereal month. During this period, the earth is moving in its orbit around the sun. The time required for the [illegible] determines the times of tide.

The tide is due [illegible] pull of the m[illegible] earth right u[illegible] site point on the other side of the earth. For this reason there are two high tides and two low tides every 24 hours and 50 min, which is the time required for the earth to turn once with reference to the moon.

The sun also exerts gravitational pull on the earth's waters and produces tidal effects, but these are much less marked than the lunar tides. When the earth, sun, and moon are in line, at new moon and full moon, the tide-producing forces operate together, making larger tides called spring tides. At the periods of first and third quarters of the moon, the sun and moon are at right angles to each other and their reduced forces produce smaller than average tides called neap tides. When spring tides coincide with the moon at perigee they are unusually high, and when neap tides coincide with the moon at apogee they are unusually small. *See* APOGEE; PERIGEE; TIDE.

[V.H.E.]

Bibliography: V. C. Finch, G. T. Trewartha, A. H. Robinson, and E. H. Hammond, *Physical Elements of Geography*, 4th ed., 1957; A. N. Strahler, *Physical Geography*, 1951.

Geography, physical

The description, analysis, classification, and genetic interpretation of the natural features and phenomena of the earth's surface.

As thus defined, physical geography is distinct from political, historical, and economic geography, regarded by many as subdivisions of human geography, as well as from plant and animal geography, frequently referred to as ecology, the study of the mutual relations between organisms and their environment. None of these geographical studies, however, can be successfully prosecuted without due consideration of the subject matter and concepts of physical geography; hence the limits of its scope cannot be established with precision.

Since the nineteenth century, as frequently occurs among the sciences, numerous specialized but overlapping disciplines have developed within the broad field of physical geography. Among these are (1) geodesy, the study and measurement of the shape and dimensions of the earth, (2) geomorphology or physiography, the interpretive description of the relief features of the earth, especially of its land forms, (3) hydrology, dealing with the water of the earth and thus embracing oceanography and limnology (the study of lakes and ponds), (4) glaciology, and (5) meteorology and climatology. Geophysics is not today a synonym for physical geography, although it has been so defined; it is rather a technique that may be applied in the study of any physical features or phenomena ranging from the center of the earth to the outermost parts of the atmosphere.

The relationship between physical geography and geology is both intimate and overlapping. Geomorphology, for example, is as much a part of the latter as of the former. In practice, no useful distinction can be drawn between geomorphology and the physical geology of Pleistocene and Recent time. When the analytical and interpretive aspects of physical geography are stressed, physical geography may be considered as the part of geology which deals with surface features of the earth as they are today.

Further considerations of physical geography appear in various topical and regional articles. For examples of principal topical discussions, *see* AERIAL PHOTOGRAPH; CARTOGRAPHY; CONTINENT; EARTH RESOURCE PATTERNS; GEOGRAPHY, MATHEMATICAL; HILL AND MOUNTAIN TERRAIN; MINERAL FUEL AREAS; MINERAL RESOURCE AREAS; PLAINS; SOIL, ZONAL DISTRIBUTION; TERRAIN AREAS, WORLDWIDE. For physical geography of the continents and major island areas, *see* AFRICA; ANTARCTICA; ARCTIC AND SUBARCTIC ISLANDS; ASIA; AUSTRALIA AND NEW ZEALAND; EAST INDIES; EUROPE; NORTH AMERICA; PACIFIC ISLANDS; SOUTH AMERICA; WEST INDIES. [K.F.M.]

Geologic thermometry

The measurement or estimation of temperatures at which geologic processes take place. Methods used can be divided into two groups, nonisotopic and

isotopic. The isotopic methods involve the determination of distribution of isotopes of the lighter elements between pairs of compounds in equilibrium at various temperatures and application of these data to problems of the temperature of formation of these compounds (commonly minerals) in nature.

NONISOTOPIC METHODS

Direct measurement. Temperatures can be measured directly in hot springs, in fumaroles, in flows of lava, and in artificial openings such as mines, boreholes, and wells.

Hot springs. The temperatures of hot springs range from slightly above the mean annual temperature of the region in which they occur to the boiling point of water at the elevation of the outlets. In other words, in temperate regions and at moderate altitudes the temperatures of hot springs range from about 20 to 100°C.

Fumaroles. At temperatures above its boiling point, water issues from vents as steam called fumaroles; temperatures of these fumaroles have been measured up to 560°C near Vesuvius, and up to 645°C in the Valley of Ten Thousand Smokes. Fumaroles in lavas may reach temperatures of 700–800°C.

Lava. Lava is molten rock coming out on the earth's surface. Its temperature on extrusion ranges from about 700–900°C for andesitic and dacitic lava to 1200°C for basaltic lava. The viscosity of lava increases with decreasing temperature and basaltic lava ceases to flow when it cools to 700–800°C. Intrusive magmas of similar compositions are probably intruded at similar temperatures, as indicated by such things as their effects on coal beds into which they are intruded, and the forms and assemblages of the first minerals to crystallize.

The temperature to which rocks around an intrusion (country rocks) are heated depends on many factors: temperature of the magma; temperature, composition, and structure of the country rock; abundance and nature of solutions given off by the intrusion; and size of the intrusion. In general, the temperature of the country rocks at the contact will be much lower than that of the magma. For example, an intrusive sheet of dolerite at 1100°C may heat the contact rock to 600–700°C.

Mines, boreholes, and wells. Temperatures have been measured in enough artificial openings in the earth's crust so that the temperature distribution is well known in many areas to depths of several thousand feet. Measurements of gradients range from about 40 to 170 ft/°C in nonvolcanic areas. Highest temperature yet encountered in such an area is 154°C from a well 20,521 ft deep in Sublette County, Wyoming.

Gradients in volcanic areas are much higher near the surface (up to 1.3 ft/°C), but at depths of 750–1000 ft a temperature of about 250°C is commonly reached and persists to much greater depths (to at least 5500 ft in Tuscany).

Calculations and extrapolations give greatly different pictures for temperature distribution from the zone of measurements to the center of the earth depending on the assumptions made. Estimates of the temperature at the center of the earth range from 1600 to 76,000°C, but most of the recent (1945–1958) estimates are in the 5000–10,000°C range. One still more recent estimate (1959) is based on experimental determination of the variation of melting point of iron with pressure up to almost 100,000 atm, and gives 2600°C as the temperature of the center of the earth and 2340°C as the minimum temperature at the boundary of the core.

Indirect methods. Some indirect nonisotopic methods appear to give estimates with accuracy of the same order of magnitude as direct measurements; others place the temperature within a certain range; still others tell only that a given process took place above or below a certain temperature.

As phase-equilibrium relationships in systems analogous to those in nature become more accurately known, it will be possible to make more accurate estimates of geologic temperatures. The relationships commonly employed are listed below.

Melting points. The melting point of a mineral, corrected for the pressure under which it was formed, gives a maximum temperature of formation for the assemblage in which it grew because other substances, especially water, lower the temperature at which a mineral will crystallize. For example, if realgar (AsS) occurs in a vein with other minerals in such a way that they must have crystallized simultaneously, then the whole assemblage must have formed at a temperature lower than 320°C, the melting point of realgar. Lists of melting points and other phase relationships important in geologic thermometry are given in E. Ingerson (1955), listed in the bibliography.

Transformation temperatures (inversions). Many minerals have two or more crystalline modifications which form, or exist, in different temperature ranges. For example, under certain conditions marcasite forms at temperatures below 300°C, but at about 450°C it transforms to pyrite at an appreciable rate. Therefore, a coprecipitated mineral assemblage including marcasite was certainly formed below 450°C and probably below 300°C.

Other pairs of minerals transform in either direction at a definite temperature; for example, low (α) quartz changes to high (β) quartz when it is heated to 573°C and high quartz changes to low when it is cooled below 573°C. Therefore, phenocrysts of high quartz in lavas were formed above 573°C, crystals of low quartz in veins, below 573°C.

Dissociation and decomposition temperatures. Many minerals break up when they are heated. If one of the products is a gas, the temperature of decomposition changes rapidly with pressure; the pressure at the time of formation must be known or estimated before such a mineral can be used as a geologic thermometer. For example, calcite dissociates into lime and carbon dioxide (CO_2) at 885°C under atmospheric pressure, but under a sufficiently high pressure of CO_2 (1025 atm) it melts at 1339°C without decomposing.

of Capricorn, 23½° south. Winter finds the overhead sun rays moving back toward the equator, which is attained on March 21 (vernal equinox). During the fall and winter, days in the Northern Hemisphere will be shorter than the nights and the sun will drop to its lowest angle above the horizon. Spring is that period when the sun is again overhead north of the Equator, from March 21 to June 21. *See* EQUINOX; SOLSTICE.

Earth-time relations. The system of time used by man is also a result of earth motions. It takes the earth 24 hours to make one complete rotation of 360°, or it rotates through 15° of longitude in 1 hour. When it is noon at a given meridian it will be 1 hour past noon 15° to the east and 1 hour before noon 15° to the west. *See* EARTH (ORBITAL MOTION); INTERNATIONAL DATE LINE; TIME.

Lunar-terrestrial relations. Although less important to man than earth-sun relationships, the moon plays a significant role as the major force in the production and timing of tides. The moon (diameter about 2160 miles) is considerably smaller than the earth but its movements are quite similar; its orbital revolution around the earth and its axial rotation are in the same direction. The distance from earth to moon varies from about 222,000 miles at its closest approach (perigee), to about 253,000 miles at its greatest distance (apogee). The moon revolves around the earth once in about 27⅓ days (27 days, 7 hours, 43 min, 11.5 sec); this is called the sidereal month. During this period, the earth is moving in its orbit around the sun. The time required for the moon to revolve around the earth and to assume the same position with respect to the sun is thus slightly longer. The period of this revolution, about 29½ days, is called the synodic month. It determines the times of tide.

The tide is due essentially to the gravitational pull of the moon; the greatest force will be on the earth right under the moon and at a directly opposite point on the other side of the earth. For this reason there are two high tides and two low tides every 24 hours and 50 min, which is the time required for the earth to turn once with reference to the moon.

The sun also exerts gravitational pull on the earth's waters and produces tidal effects, but these are much less marked than the lunar tides. When the earth, sun, and moon are in line, at new moon and full moon, the tide-producing forces operate together, making larger tides called spring tides. At the periods of first and third quarters of the moon, the sun and moon are at right angles to each other and their reduced forces produce smaller than average tides called neap tides. When spring tides coincide with the moon at perigee they are unusually high, and when neap tides coincide with the moon at apogee they are unusually small. *See* APOGEE; PERIGEE; TIDE.

[V.H.E.]

Bibliography: V. C. Finch, G. T. Trewartha, A. H. Robinson, and E. H. Hammond, *Physical Elements of Geography*, 4th ed., 1957; A. N. Strahler, *Physical Geography*, 1951.

Geography, physical

The description, analysis, classification, and genetic interpretation of the natural features and phenomena of the earth's surface.

As thus defined, physical geography is distinct from political, historical, and economic geography, regarded by many as subdivisions of human geography, as well as from plant and animal geography, frequently referred to as ecology, the study of the mutual relations between organisms and their environment. None of these geographical studies, however, can be successfully prosecuted without due consideration of the subject matter and concepts of physical geography; hence the limits of its scope cannot be established with precision.

Since the nineteenth century, as frequently occurs among the sciences, numerous specialized but overlapping disciplines have developed within the broad field of physical geography. Among these are (1) geodesy, the study and measurement of the shape and dimensions of the earth, (2) geomorphology or physiography, the interpretive description of the relief features of the earth, especially of its land forms, (3) hydrology, dealing with the water of the earth and thus embracing oceanography and limnology (the study of lakes and ponds), (4) glaciology, and (5) meteorology and climatology. Geophysics is not today a synonym for physical geography, although it has been so defined; it is rather a technique that may be applied in the study of any physical features or phenomena ranging from the center of the earth to the outermost parts of the atmosphere.

The relationship between physical geography and geology is both intimate and overlapping. Geomorphology, for example, is as much a part of the latter as of the former. In practice, no useful distinction can be drawn between geomorphology and the physical geology of Pleistocene and Recent time. When the analytical and interpretive aspects of physical geography are stressed, physical geography may be considered as the part of geology which deals with surface features of the earth as they are today.

Further considerations of physical geography appear in various topical and regional articles. For examples of principal topical discussions, *see* AERIAL PHOTOGRAPH; CARTOGRAPHY; CONTINENT; EARTH RESOURCE PATTERNS; GEOGRAPHY, MATHEMATICAL; HILL AND MOUNTAIN TERRAIN; MINERAL FUEL AREAS; MINERAL RESOURCE AREAS; PLAINS; SOIL, ZONAL DISTRIBUTION; TERRAIN AREAS, WORLDWIDE. For physical geography of the continents and major island areas, *see* AFRICA; ANTARCTICA; ARCTIC AND SUBARCTIC ISLANDS; ASIA; AUSTRALIA AND NEW ZEALAND; EAST INDIES; EUROPE; NORTH AMERICA; PACIFIC ISLANDS; SOUTH AMERICA; WEST INDIES.

[K.F.M.]

Geologic thermometry

The measurement or estimation of temperatures at which geologic processes take place. Methods used can be divided into two groups, nonisotopic and

assemblage indicates temperature of formation not over 200°C and possibly below 100°C. The lower limit of formation of these minerals is not known.

Similar experiments give analogous results for other minerals and mineral groups. Sericite, for example, forms at about 200–525°C in slightly basic to somewhat more acid solutions if aluminum and potassium are both high. Pyrophyllite forms at about 300–550°C if aluminum and potassium are both low. Serpentine cannot form above about 500°C even under very high pressures of water vapor, and most varieties of chlorite crystallize below 500°C. Muscovite is stable from about 400 to 800°C at pressures likely to be involved in rocks formed at depths small enough so they can be brought to the surface by erosion. Phlogopite forms from about 800 to 1100°C in the same pressure range. Talc does not form above about 825°C.

Where two or more minerals were formed in equilibrium it may be possible to narrow the temperature range considerably. Although talc is stable at a temperature up to 825°C, in equilibrium with enstatite at moderate pressures the assemblage is stable only from about 670 to 800°C. Likewise, serpentine can form in equilibrium with brucite only below 450°C at moderate pressures or below 400°C at low pressures.

Similar relationships have been established for hydrous aluminum oxides and silicates, carbonates, and evaporites (E. Ingerson, 1955).

Other indirect methods. Other methods commonly used in geologic thermometry are discussed in this section.

Fossil assemblages. By determining the temperatures of the water in which certain types of organisms grow, it is possible to infer the temperature of the water at the time strata containing fossils of the same species, or perhaps closely similar species, were laid down. Thus far, this method has been confined to differentiation between cold- and warm-water assemblages, but as more information is obtained about temperatures at which certain species lived it should be possible to assign temperature ranges to the water in which the formations containing them were laid down.

Properties dissipated by heating. Properties such as thermoluminescence, radiation colors, and metamictization, which are exhibited by many minerals (and thermoluminescence by some rocks), are dissipated by heating. Most thermoluminescence is dissipated below 250°C, although a few materials have been reported which retain some capacity for thermoluminescence up to about 500°C. Likewise, most radiation colors in minerals are dissipated at temperatures below 300°C, but in a few they persist up to 500°C or higher.

Metamictization is the destruction of the regular internal structure of a mineral produced by emanations from contained radioactive elements. The metamictization of minerals can be dissipated and the original structure restored by heating them to about 450–900°C, depending upon the mineral and the rate and time of heating.

Possession of these properties by minerals does not mean that they were formed at temperatures lower than those at which the properties are dissipated, but that they have not been heated to higher temperatures since the properties were acquired. In deducing the thermal history of such materials, therefore, the problem of when the property was acquired becomes important.

Crystallography. The generalization has been made that crystals grown at relatively low temperatures are likely to be simple in habit; those grown at high temperatures, complex. The composition and pH of the solution, presence of impurities, rate of growth, and other factors can affect crystal habit.

Potassium feldspar is a good example of the change of crystal habit with temperature of formation. Phenocrysts of potassium feldspar in porphyries (800°C±) are dominated by base, clinopinacoid, and orthodomes, giving crystals elongate parallel to the *a* axis. Crystals in pegmatites (500°C±) are elongate parallel to the *c* axis and are dominated by (010), (001), and (110). In high-temperature veins such as those of the Alps (350°C±) the (110) form becomes more prominent and the crystals are simpler. In very low-temperature veins the potassium feldspar is adularia (150°C±) and only the two forms (110) and (101) remain, so the crystals are as simple as possible. Other examples are quartz, calcite, and fluorite (Ingerson, 1955).

Liquid inclusions. Minerals crystallizing from aqueous solutions commonly have imperfections that retain samples of the solution as liquid inclusions in the final crystals. When the crystal cools the solution contracts and a vapor bubble appears in each liquid inclusion. By heating plates of the mineral on a heating stage on a microscope and determining the temperature at which the solution just fills the cavities it is possible to estimate temperature of formation, if the pressure at formation was essentially the same as the vapor pressure of the solution. For significantly higher pressures it is necessary to estimate what the pressure was and apply a correction.

The following necessary assumptions appear to be justified in most, if not all, cases. (1) The inclusion cavities were just filled with fluid under the temperature and pressure prevailing during crystallization. (2) Change of volume of the mineral itself is not significant. (3) Changes in volume and concentration brought about by deposition of material during cooling are such as not to affect the result. (4) Primary and secondary liquid inclusions can be distinguished under the microscope. (5) There has been no leakage from or into the inclusions. (6) The liquid is an aqueous solution containing no carbon dioxide or other gas in large concentration. (7) Pressure-volume-temperature relations are near enough to those of pure water or chloride solutions that have been studied so that no seriou errors are introduced by using data that are a able.

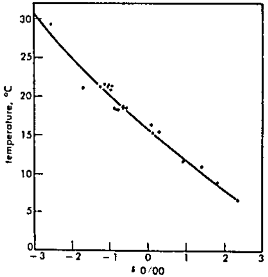


Fig. 1. Isotope temperature scale. (*From S. Epstein, R. Buchsbaum, H. A. Lowenstam, and H. C. Urey, Bull. Geol. Soc. Am., 64(11):1315–1326, 1953*)

Temperature ranges that have been estimated by this method for some common vein and pegmatite minerals are as follows: calcite, 40–362°C; sphalerite, 75–275°C; fluorite, 83–350°C; vein quartz, 100–440°C; pegmatite quartz, 200–530°C; and topaz, 275–500°C. [E.I.]

Bibliography: E. Ingerson, Geologic thermometry, *Crust of the Earth*, Geol. Soc. Am. Spec. Paper 62, 1955. E. Ingerson, Methods and problems of geologic thermometry, *Econ. Geol.*, 50th anniv. vol., pp. 341–410, 1955.

ISOTOPIC METHODS

In many cases the ratio of the stable isotopes of the lighter elements is not uniform in nature. The nonuniformity in the isotope ratio is a result of many natural processes and the temperatures at which these processes occur. Because temperature is a factor in the determination of isotope distribution, the isotope distribution provides information about the temperature conditions at which certain natural processes occur. For example, the knowledge of the isotopic composition of oxygen of minerals in certain rocks can give information about the temperature at which the rock was formed. Thus, measurements of relative isotope abundance can be useful in geologic thermometry.

Principles. The method of measuring temperature by the isotopic analysis is based on the fact that the chemical and physical properties of a chemical compound depend not only upon the elements which form it but also upon the isotopic composition of the elements from which it is constituted. For example, the standard free energy of formation for $C^{12}O^{16}$ is not identical with that of $C^{13}O^{16}$. The differences in properties between the two molecules are sufficiently small so that the natural variations in the isotopic composition of the elements can be ignored (for all practical purposes) when dealing with the usual chemical problems. On the other hand, these small differences can be utilized in studying many problems, among them the determination of temperature at which many natural processes take place.

For a specific illustration, consider a system at some temperature composed of carbon dioxide (CO_2) and liquid water (H_2O). The two components are in chemical and isotopic equilibrium, that is, the isotopes of the element common to both components have distributed themselves between CO_2 and H_2O so that the system is at a minimum free energy. Such a system can be represented by an exchange reaction

$$CO_2^{16}(g) + 2H_2O^{18}(l) = CO_2^{18}(g) + 2H_2O^{16}(l) \qquad (1)$$

The equilibrium constant K for such a reaction is

$$K = \frac{[CO_2^{18}]/[CO_2^{16}]}{[H_2O^{18}]^2/[H_2O^{16}]^2} \qquad (2)$$

It can be readily shown that

$$K^{1/2} = \frac{[O^{18}]/[O^{16}] \quad (\text{in the } CO_2)}{[O^{18}]/[O^{16}] \quad (\text{in the } H_2O)} \qquad (3)$$

where the $[O^{18}]/[O^{16}]$ ratio in the carbon dioxide is equal to

$$\frac{2[CO_2^{18}] + [CO^{16}O^{18}]}{[CO^{16}O^{18}] + 2[CO_2^{16}]}$$

The ratio of ratios given in Eq. (3) is referred to as the fractionation factor α. If the standard free energies of formation of CO_2^{16} and CO_2^{18} were identical and the same were true for H_2O^{16} and H_2O^{18}, the standard free energy change ($\Delta F°$) for the reaction (1) would be zero, and from the well-known thermodynamic equation

$$\Delta F° = -RT \ln K \qquad (4)$$

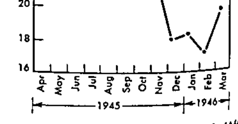


Fig 2. Sea-surface temperatures in Ferry Reach. (*After H. B. Moore, S. Epstein and H. A. Lowenstam, J. Geol., 61:424–438, 1953*)

K and α would be equal to unity. However, $\Delta F°$ is not equal to zero, and $\alpha = 1.04$ at 25°C.

Equation (4) also shows that α changes with temperature. A knowledge of the relationship between α (or the $[O^{18}]/[O^{16}]$ ratios of the two components) and temperature will then permit the determination of temperatures from the measurements of the isotopic composition of oxygen. Isotope exchange reactions have been considered for many systems involving oxygen compounds as well as a variety of other elements and their isotopes such as H_2, B, C, and N. In principle, for each of these systems, an α can be evaluated and for every α a potential thermometer can exist. In some cases the fractionation factors are calculable by the methods of statistical mechanics, using spectroscopic data. This was done by H. C. Urey and his students as early as 1934.

First application. The first isotope thermometer was developed by a group headed by Urey in 1951. This thermometer was based on the calcium carbonate–water system, in which the change with temperature of

$$\alpha = \frac{[O^{18}]/[O^{16}]_{\text{carbonate}}}{[O^{18}]/[O^{16}]_{H_2O}}$$

is used for the thermometer. In this case the thermometer is used primarily for determining temperatures of localities at which marine animals deposit their calcareous skeleton. Under these conditions of deposition, the amount of calcium carbonate is very small compared to the amount of water with which the carbonate is in equilibrium. Any change in the fractionation factor will, therefore, affect the $[O^{18}]/[O^{16}]$ ratio of the carbonate only. If, for example, α decreased as a result of increase in temperature, the O^{18} in the carbonate will decrease to cause the $[O^{18}]/[O^{16}]$ ratio to decrease. The quantity of O^{18} atoms transferred to the water will be so small compared to the O^{18} present in the water that the $[O^{18}]/[O^{16}]$ ratio of the water will not be significantly changed. It follows therefore that temperature will affect only the $[O^{18}]/[O^{16}]$ ratio of the skeleton material.

Experimental procedure. The techniques involved can be briefly summarized. A Nier-type mass spectrometer was modified to permit the measurements of the $[O^{18}]/[O^{16}]$ ratio to sufficient accuracy.

The $[O^{18}]/[O^{16}]$ ratio is difficult to measure to a sufficiently high accuracy. Because it is of interest to know the change of ratio from sample to sample, it is necessary to measure this difference accurately. This can be done with the necessary precision by the use of the modified mass spectrometer. The isotope data is reported in terms of change in δ per mil of the ratio relative to a standard gas

$$\delta = \frac{[O^{18}]/[O^{16}]_{(\text{sample})} - [O^{18}]/[O^{16}]_{(\text{standard})}}{[O^{18}]/[O^{16}]_{(\text{standard})}} \times 1000$$

where δ equals 10 means the $[O^{18}]/[O^{16}]$ ratio of the sample is 10 per mil or 1% greater than that of the standard.

The term δ equals -10 means that the ratio is smaller than that of the standard by 10 per mil or 1%.

Carbon dioxide from calcium carbonate precipitated under equilibrium conditions at 16.5°C has been used as a standard. Carbon dioxide is the gas used in the mass spectrometer for the isotope measurements. For the present purposes, the ratio of mass 46 ($C^{12}O^{16}O^{18}$) to mass 44 ($C^{12}O^{16}O^{16}$) gives the $[O^{18}]/[O^{16}]$ ratio. Hence δ can be measured to an accuracy of 0.1 per mil. With this accuracy, a change in the fractionation factor caused by 1C° change could be measured.

The chemical techniques required to extract the oxygen (in form of CO_2) from the carbonates involve the acidification of the calcium carbonate with 100% phosphoric acid. If the acidification procedure is the same for all the samples, variation in the $[O^{18}]/[O^{16}]$ ratio of the extracted CO_2 is the same as that in the calcium carbonate.

The oxygen isotopic composition of the water in which the calcium carbonate grows is measured by equilibrating a small quantity of carbon dioxide with the water and then analyzing the resulting carbon dioxide. Carbon dioxide equilibrated in this way with mean ocean water at 25.3°C gives a δ value of zero.

The relationship between temperature and the value of CO_2 from calcareous skeletons grown in mean ocean water is shown in Fig. 1. The curve is represented by the relationship

$$T = 16.5 - 4.3 \times \delta_{\text{carbonate}}$$

The various points on the curve were obtained by analyzing calcareous skeletons, either grown in tanks under controlled conditions or found in natural habitats of known temperature.

A present-day assemblage of marine animals living off the coast of Bermuda was studied for the growth temperature of the skeletons. The temperature records of the localities were well known. The temperatures ranged from 18 to 29°C (Fig. 2). These temperatures, recorded by the isotopes of the shell material compared to the temperature of the water, provided valuable information in determining the reliability of isotope temperatures recorded by marine calcareous skeletons. Figure 3 shows the temperature recorded by the skeleton of a marine snail (*Strombus gigas*). The data in Fig. 3*a* were obtained by analyzing successive small increments of shell in the sequence laid down by the animal. Each increment represents a short growth period. These animals appear to lay down shells practically the year around. The equivalent fossil shell which grew some hundreds of thousands of years ago shows a similar record (Fig. 3*b*). The pelecypod analyzed (*Chama acrophylla*) retains the shell grown only during a narrow temperature range. Information of this type is r… to interpretation of fossil material, be…

curate temperature records must be interpreted with realization that certain skeletons will not record the complete range of temperatures existing in a habitat. The data on recent shells verify the validity of the isotope temperature method and indicate some possible biological studies associated with marine animal growth.

Paleotemperatures in the ocean. The determination of ocean temperatures using the isotope thermometer of fossil material has several difficulties not inherent in studies of present-day temperatures. The fossil calcium carbonate skeleton must be one that was originally laid down, and laid down in isotopic equilibrium. The isotopic composition of the oxygen of the calcium carbonate depends not only upon the temperature, but also upon the $[O^{18}]/[O^{16}]$ ratio of the ocean water. Present-day ocean water can be analyzed, but fossil ocean water is not available. The first problem is the less serious one because it is possible to recognize original material by the structure of the crystals of the calcium carbonate.

Skeletal remains originating as far back as the Paleozoic era (older than 200,000,000 years) have been found with their original carbonate preserved. Younger fossil remains have a higher probability of preservation, so that a large time span of the earth's climatic history is available for the application of the isotope thermometer. *See* PALEOECOLOGY (GEOCHEMICAL ASPECTS).

The second problem, involving the uniformity of the $[O^{18}]/[O^{16}]$ ratio of the ocean waters, is somewhat more serious. If the $[O^{18}]/[O^{16}]$ variations in the oceans of the past were similar to those of the present oceans, large errors in isotope temperatures could be made. The $[O^{18}]/[O^{16}]$ ratio of the present oceans varies sufficiently to cause errors of temperatures as large as 10°C. The variations in the $[O^{18}]/[O^{16}]$ ratio of the present oceans appear to be due primarily to the existence of glaciers, which are a source of water of very low $[O^{18}]/[O^{16}]$ ratio. Melt water is added to the cold currents of the oceans, and water vapor of low $[O^{18}]/[O^{16}]$ ratio is removed from the warmer ocean surfaces, causing heterogeneity in the isotope ratios of the oceans. Because there is little evidence that glaciers existed during the large portion of the earth's post-Paleozoic history, it can be surmised that the $[O^{18}]/[O^{16}]$ ratio variation of the oceans of the past was markedly less and that errors of only a few degrees are introduced in the isotope thermometer. It can be expected that large errors would occur in habitats near mouths of rivers or areas of very high evaporation, but these would be anomalous areas and easily recognizable from other evidence. The problem of the variations of the $[O^{18}]/[O^{16}]$ ratio of ocean water could be entirely eliminated if another fossil oxygen compound could be used in association with calcium carbonate, so that the thermometer could be based on the fractionation factor between the two available compounds. The $[O^{18}]/[O^{16}]$ ratio of the ocean water of the past would then not enter into the temperature determination. The possibility of using calcium phosphate and calcium carbonate as a pair has been considered, but has not as yet been developed.

The first application of the carbonate temperature scale to fossil material was the determination of temperatures from a Jurassic belemnite skeleton. This skeleton is cigar-shaped and is composed of solid calcite. It was internally grown by the animal, laying down layer upon layer of calcite, so that a cross section of the skeleton has a pattern similar to the ring pattern of a cross section of a tree trunk. Successive layers of a belemnite skeleton were ground off and each increment of powder was analyzed for the $[O^{18}]/[O^{16}]$ ratio. Each increment represented about one-tenth of one year's growth.

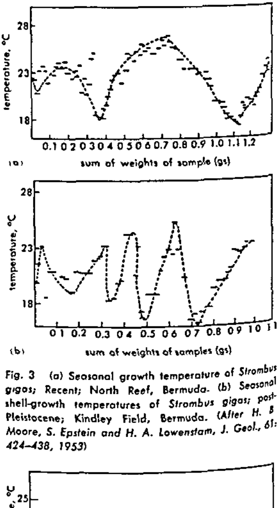


Fig. 3 (a) Seasonal growth temperature of *Strombus gigas;* Recent; North Reef, Bermuda. (b) Seasonal shell-growth temperatures of *Strombus gigas;* post-Pleistocene; Kindley Field, Bermuda. *(After H. B. Moore, S. Epstein and H. A. Lowenstam, J. Geol., 61: 424–438, 1953)*

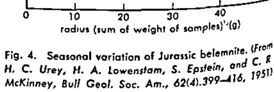


Fig. 4. Seasonal variation of Jurassic belemnite. *(From H. C. Urey, H. A. Lowenstam, S. Epstein, and C. R. McKinney, Bull Geol. Soc. Am., 62(4).399–416, 1951)*

Figure 4 shows the temperature recorded by successive increments of skeleton. It can be seen that a seasonal variation in the temperature is recorded and that the animal lived about 4 years. Had the original skeleton been modified while buried, then it would be expected that the variation found here would be obliterated. These results gave excellent evidence for the feasibility of applying the temperature scale, even as far back as the Jurassic times.

This work was extended to include the study of the Upper Cretaceous period and some of the temperatures determined are shown in Fig. 5.

The data for Fig. 5 were obtained from pieces of belemnite and brachiopod skeletons from the Upper Cretaceous period (90,000,000–120,000,000 years ago). Each datum represents the average yearly growth temperature. The samples were collected from a number of localities in Sweden, Denmark, Holland, England, France, and the southeastern United States. The temperature record of the samples indicates that the Upper Cretaceous period was, on the whole, warmer than the present, a fact well known by geologists from other lines of evidence. In addition, it appears that the early and later parts of the Upper Cretaceous period were cooler than the middle part of the period. The lowering of the temperature as the period approached its end is of interest because it was at the end of the Upper Cretaceous that a number of drastic changes took place in the evolution of life, including the extinction of several important forms of life, such as the dinosaurs. Any theory dealing with these evolutionary aspects must be compatible with the climatic record.

Pleistocene climate. An interesting application of the isotope thermometer was made on relatively young skeletal remains. These are some of the Foraminifera, small calcareous Protozoa, which constitute a large fraction of ocean bottom sediments in some of the oceanic areas. These skeletons, of the species which float near the surface when alive, should preserve a record of surface temperature of open oceans, starting from the present, and going back in time. Climatically, the last few hundred thousand years were very interesting. Field evidence indicates that there were a number of great temperature fluctuations, resulting in several blankets of ice over more than one-half of the North American continent. Results are shown in Fig. 6 and indeed there is an excellent and continuous record of reasonable temperature in the Foraminifera tests. The maximum southward ice advances, designated by the name ice ages, are accompanied by a record of minimum temperatures. *See* MARINE SEDIMENTS.

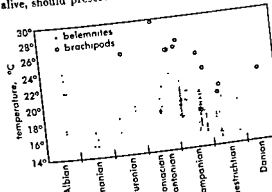


Fig 5. Mean-temperature distribution of belemnites and brachiopods from Albian to Danian in western Europe. (From H. A. Lowenstam and S. Epstein, *J. Geol.*, 62(3):207–248, 1954)

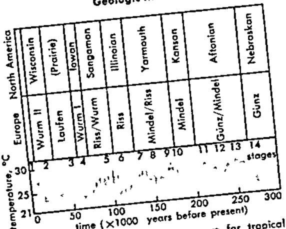


Fig. 6. Generalized temperature curve for tropical surface waters and continental correlations. (From C. Emiliani, *J. Geol.*, 66(3):264–275, 1958)

The above examples of the application of the carbonate isotope thermometer are but a few of many interesting researches dealing with climatic temperatures that are of geological interest.

Temperature in geologic environment. The most recent attempts to apply isotope studies to geologic thermometry deal with problems of the conditions under which common rock and economically interesting rocks are formed.

In this connection, the relationships of the $[O^{18}]/[O^{16}]$ ratio among minerals in a rock are temperature-dependent. An exact thermometer for such systems has, as yet, not been developed. Neither is it, as yet, certain that relatively precise isotope thermometers are possible for rocks formed at relatively high temperatures. It is first necessary to calibrate an isotope thermometer by growing minerals in equilibrium with one another at known higher temperatures, and by measuring the $[O^{18}]/[O^{16}]$ ratio of these minerals. This has not, as yet, been done. However, temperature trends have been determined from isotope studies of certain coexisting minerals which show excellent correlation with field evidence and other chemical considerations. These preliminary isotope investigations were necessary to determine whether or not the rock-forming minerals contain an isotope record w may be interpretable in terms of temperat other factors.

e thermometry aspect of the probl
zing the $[O^{18}]/[O^{16}]$ ratio

pairs, calcite-quartz and quartz-magnetite. The samples chosen represent a wide range of temperature of formation as indicated by other well-known geological evidence.

The chemical techniques to extract the oxygen from minerals other than calcite require high-temperature reactions with either graphite or fluorine. In the former case, carbon monoxide is released and then converted to carbon dioxide by means of a nickel catalyst. In the fluorine reaction oxygen is released. The oxygen is converted to carbon dioxide by treating it with graphite. In this way both techniques permit the conversion of the oxygen of the mineral to carbon dioxide, which is the most convenient gas to use in a mass spectrometer for analysis of the $[O^{18}]/[O^{16}]$ ratio.

The results of some of the oxygen isotope researches pertaining to isotope thermometry are shown in the accompanying table. The mineral pairs are listed in order of increasing temperature of formation. The nature of the deposit from which the mineral pairs were collected is also designated. The differences in δ values between the mineral pairs (designated by a Δ) decrease with increased temperature; that is, as would be predicted from theoretical considerations, the $[O^{18}]/[O^{16}]$ ratios of the coexisting mineral pairs become more similar with increasing temperature. The distribution of the $[O^{18}]/[O^{16}]$ ratio among the minerals indicates that there is a good possibility that isotope thermometers may be used to

The δ values of cogenetic quartz and calcites

Description	Mineral	δ, ‰	Δ, ‰ quartz-calcite, ‰
Hard black chert (microcrystalline quartz) in a white calcareous chalk; Cretaceous age	Quartz Calcite	31.1 26 7	7.4
Black chert in fine-grained limestone; Mississippian age	Quartz Calcite	29 2 22 6	6 6
Herkimer diamonds (quartz crystals in cavities in sandstone), small calcite crystals partly embedded in the quartz	Quartz Calcite	23 8 18.5	5 3
Recrystallized chert nodule in calcite marble; wollastonite surrounds the chert	Quartz Calcite	23 8 19 6	4 2
Chert from same bed as sample 2, but here limestone recrystallized to coarse calcite	Quartz Calcite	22.1 17.6	4 5
Hydrothermal vein containing coarsely crystalline intergrown quartz and calcite	Quartz Calcite	18 6 15.2	3 4
Hydrothermal zinc ore containing sphalerite, chalcopyrite, calcite, and quartz	Quartz Calcite	17.0 12.5	4 5
High-temperature hematite-calcite-quartz occurrence	Quartz Calcite	10 4 7.4	3.0
Quartz and calcite from core of a pegmatite	Quartz Calcite	11.2 8.6	2 6
Scheelite ore containing quartz and calcite	Quartz Calcite	11 9 10 0	1.9

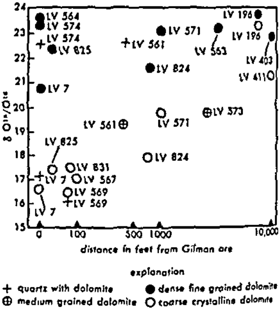


Fig. 7. The relationship between texture, oxygen isotope composition, and distance from ore body of samples of hydrothermal dolomite collected in the Gilman area, Colorado. The dense, fine-grained dolomite is practically unrecrystallized. (From A. E. J Engel, R. N. Clayton, and S. Epstein, J. Geol., 66(4): 374–393, 1958)

measure temperatures at which many minerals and rocks have been formed. An application of isotope thermometry has been made to a mining locality to determine whether the isotope ratios vary with proximity to an ore body. In the Leadville, Colorado area, sulfide deposits exist where it is thought that their formation is associated with hydrothermal activity. Around this deposit an extensive halo of recrystallized carbonate has been formed. It can be surmised that the carbonate and quartz, recrystallized near the ore, were deposited at the highest temperatures. The temperature of recrystallization should decrease as the distance from the ore body increases. A series of samples, dolomite ($MgCaCO_3$), calcite, and quartz were analyzed for the $[O^{18}]/[O^{16}]$ ratios. As can be seen from Fig. 7, the farther the samples are from the ore body, the greater is the difference in the $[O^{18}]/[O^{16}]$ ratios between the quartz and calcite. It should be possible, from this technique, to localize hot spots and determine their economic merit.

The utilization of the isotope study to problems in thermometry is, as yet, in its initial stages and its potential usefulness awaits a great deal of research. [S.EN.]

Bibliography: P. H. Abelson (ed.), *Researches in Geochemistry*, 1959; R. N. Clayton and S. Epstein, The relationship between O^{18}/O^{16} ratios in coexisting quartz, carbonate, and iron oxides from various geological deposits, *J. Geol.*, 66(4):352–373, 1958; C. Emiliani, Paleotemperature analysis of core 280 and Pleistocene correlations, *J. Geol.*, 66(3):264–275, 1958; A. E. J. Engel, R. N. Clayton,

and S. Epstein, Variations in isotopic composition of oxygen and carbon in Leadville limestone (Mississippian, Colorado) and in its hydrothermal and metamorphic phases, *J. Geol.*, 66(4):374–393, 1958; S. Epstein, R. Buchsbaum, H. A. Lowenstam, and H. C. Urey, Revised carbonate-water isotopic temperature scale, *Bull. Geol. Soc. Am.*, 64(11): 1315–1326, 1953; S. Epstein and T. Mayeda, Variation of O^{18} content of waters from natural sources, *Geochim. et Cosmochim. Acta*, 4:213–224, 1953; H. A. Lowenstam and S. Epstein, Paleotemperatures of the post-Aptian Cretaceous as determined by the oxygen isotope method, *J. Geol.*, 62(3): 207–248, 1954; J. M. McCrea, On the isotopic chemistry of carbonates and a paleotemperature scale, *J. Chem. Phys.*, 18(6):849–857, 1950; H. C. Urey, H. A. Lowenstam, S. Epstein, and C. R. McKinney, Measurement of paleotemperatures and temperatures of the Upper Cretaceous of England, Denmark, and the southeastern United States, *Bull. Geol. Soc. Am.*, 62(4):399–416, 1951.

Geological time scale

The assumed absolute time intervals occupied by the various geological periods. The evolution of plants and animals and the sequence of depositional and orogenic events provide the data of earth history. In order to relate these events to absolute time and therefore establish a geological time scale, an independent means of dating rocks is required. Before the discovery of radioactivity, crude estimates of the length of geologic history were made based on the total thicknesses of sedimentary rock and assumed rates of erosion and sedimentation. These estimates varied by as much as a factor of ten.

Development of quantitative scale. With the development of isotopic methods of determining the absolute age of rocks and minerals a quantitative geological time scale became possible. *See* ROCK (AGE DETERMINATION). From the discovery of radioactivity in 1893 until 1947 only the decay of uranium to lead was employed on macroscopic uranium minerals. Using the available data which consisted of relatively coarse measurements on a few samples, A. Holmes proposed the first quantitative time scale in 1913. His revised edition, published in 1947, was based essentially on five localities from which uranium minerals had been analyzed isotopically by A. O. Nier. Interpolation between these points was done on the assumption that the length of a geologic period is proportional to the maximum thickness of sedimentary rocks formed during that period. Subsequent work has shown that in two cases (Bedford, N.Y., and Middletown, Conn.) the stratigraphic assignment is uncertain and the pegmatites at these localities were produced in younger orogenic events than Holmes had assumed. A recent comprehensive study of the Swedish black shale containing mid-Upper Cambrian fossils has shown that it must be at least 60,000,000 years older than the preliminary age used by Holmes. Further, the Joachimsthal pitchblende point suffers from uncertainties in the absolute age. Despite these serious defects, it is a tribute to Holmes' ingenuity and the high quality of Nier's analytical work that the modification of this scale, which has been used by geologists for three decades, is not too great.

Principles. The modern geologic time scale is based on many measurements of various rock types by the three quantitative isotopic chronometers

Table 1. Definitive points in the geologic time scale

Locality	Method of determination	Stratigraphic position	Age, 10^6 years
Central City, Colo.	U-Pb (pitchblende veins)	Late Paleocene (post-Fort Union, pre-Eocene)	61 ± 5
Caucasus, U.S.S.R.	K-A (granite)	Cuts upper Eocene, overlain by lower Oligocene	40 ± 5
Coast Range, Calif.	K-A (granites)	Mid-Upper Cretaceous	80 ± 4
Sierra Nevada	K-A (granites)	Late Upper Jurassic (cuts Kimmeridgian Stage)	139 ± 4
Kelasury, U.S.S.R.	K-A (granite)	Cuts Middle Jurassic, overlain by Upper Jurassic	161 ± 5
Solikamsk, U.S.S R.	K-Ca (sylvite)	Lower middle Permian	211 ± 7
Dartmoor, England	K-A (granite)	Lowermost Permian (post-Westphalian, pre-Middle Permian)	275 ± 10
Vosges, France; Schwarzwald, Germany	K-A, Rb-Sr (granite)	Lower Carboniferous (pre-Visean, post-Dinantian)	330 ± 10
Chattanooga, Tenn.	U-Pb (black shale)	Devonian-Carboniferous boundary	350 ± 10
Shap, England	K-A (granite)	Post-Silurian (probably Lower Devonian)	390 ± 10
Maine	K-A, Rb-Sr (granite)	Post-Silurian, pre-Upper Devonian	390 ± 15
Vastergotland, Sweden	U-Pb (black shale)	Mid-Upper Cambrian	500 ± 10 (min)
Wichita Mts., Oklahoma	K-A, Rb-Sr, U-Pb (granite)	Pre-Upper Cambrian	520 ± 20

(uranium-lead, U-Pb; rubidium-strontium, Rb-Sr; and potassium-argon, K-Ar). Unfortunately it is not yet possible to measure directly the absolute age of a sedimentary rock. The scale is therefore built up from such data as the following.

1. Isotopic age measurement, generally by the K-Ar and Rb-Sr method on mica from granites which intrude fossiliferous strata. These measurements provide firm minimum ages for the intersected strata. If the granite is also overlain by later fossiliferous rocks, a maximum age for these rocks is attained. If the time interval between the fossiliferous rocks is sufficiently short, the granite provides a basic point in the absolute time scale. In those rare cases where a lava flow containing mica is intercalated in a continuous sedimentary sequence, the ideal is reached.

2. Similar measurements on mica from metamorphosed sedimentary rocks that either were covered shortly thereafter by fossil-bearing sedimentary rocks or whose age of metamorphism can be inferred from general geologic consideration.

3. K-Ar measurements on glauconite or effusive rocks enclosed in a sedimentary sequence. Because of the uncertain retention of argon these data provide only firm minimum ages for the strata tested.

4. U-Pb measurements on uranium-rich black shales. Depending on the situation, these may provide actual or minimum ages for the stratigraphic unit involved.

Table 1 lists a few of the most definitive points available in 1959 for the construction of the time scale. The data on glauconite and effusive rocks are not included but a reasonably complete list is given in J. L. Kulp, Absolute age determination of sedimentary rocks, *World Petrol. Congr. Proc. 5th Congr., New York*, 1959.

The new time scale. A new geological time scale using all of these data has been constructed. It is shown in Table 2 along with Holmes' revised (1947) scale. The over-all effect of the new scale is an extension of about 12% beyond the Cretaceous Period. In the new scale the Jurassic, Permian, and Carboniferous Periods are approximately doubled in length. The base of the Upper Cambrian appears to date back about 510,000,000 years. The base of the Cambrian cannot be precisely defined by paleontological methods at present. In many areas of the world there appears to be a more or less continuous sedimentary sequence extending far below the lowest Cambrian fossil zone (*Olenellus*). With current information it might be best to base the absolute paleontological scale on the well-defined boundary between the Lower Cambrian and the Middle Cambrian. The base of the Cambrian may either be set arbitrarily at about 600,000,000 years ago, which would appear to cover the possible range of *Olenellus*, or be held in abeyance pending the discovery of suitable pre-*Olenellus* fossils. *See* CAMBRIAN.

Reliable dates for subdividing the later Cenozoic are still scarce. The dating of the beginning of the Recent at 11,000 years ago is based on the major

Table 2. Geological time scale

Era / Period / Epoch	Age of beginning of period × 10^6 years: Holmes 1947 B scale	Age of beginning of period × 10^6 years: New scale	Length of period, × 10^6 years
Cenozoic			
Quaternary			
Recent		0 011	
Pleistocene	1	0 6	
Tertiary			65
Pliocene	12	12	
Miocene	26	20	
Oligocene	38	35	
Eocene		55	
Paleocene		65	
Mesozoic			
Cretaceous			75
Upper (Base Santonian)		90	
Middle (Base Albian)		120	
Lower	127	140	
Jurassic			45
Upper (Base Callovian)		155	
Middle (Base Bajocian)		170	
Lower	152	185	
Triassic			45
Upper		200	
Middle		215	
Lower	182	230	
Paleozoic			
Permian			45
Upper (Base Ochoan)		245	
Middle (Base Guadalupian)		260	
Lower	203	275	
Pennsylvanian			35
Upper			
Middle			
Lower		310	
Mississippian			40
Upper			
Middle			
Lower	255	350	
Devonian			50
Upper		365	
Middle		385	
Lower	313	400	
Silurian			20
Upper			
Middle			
Lower	350	420	
Ordovician			70
Upper		440	
Middle		460	
Lower	430	490	
Cambrian			50+
Upper (Base Croixian)	450	510	
Middle		540	
Lower			
? ?			
Precambrian			
Sinian sedimentation, U.S.S.R.			600–1200
Grenville orogeny			900–1200
Central and Western United States basement			1300–1400
Rapakivi, Finland, and U.S.S.R.			1600
Sveco-Fennidic orogeny			1800–1900
Central Canadian Shield, Western Australia, Ukraine			2500–2800
Oldest reported basement, Kola Peninsula, U.S.S.R.			~3400
Probable age of the earth			~4600

change in world climate at that time. If the beginning of the Pleistocene is defined as the time of the first continental glaciation, new work suggests that 600,000 years ago would be a more reasonable estimate than 1,000,000 years, which has been in general use.

Below the Cambrian and thus in the absence of suitable fossils, it is not possible to identify world-wide events independently of absolute age measurements. Instead, regional orogenies or local sedimentation must be defined in terms of the absolute time scale. The oldest geologic event detected as of 1959 appears to be an orogeny in the Kola Peninsula of the U.S.S.R. where micas give K-Ar ages of about 3,500,000,000 years. The age of the earth is probably about 4,600,000,000 years; thus, one-quarter of earth history remains to be discovered. See PRECAMBRIAN.

The new time scale presented in Table 2 will undergo continual revision and refinement as intensive research on these problems continues. It appears that many stratigraphic horizons will eventually be dated by isotopic methods so that the amount of interpolation will be greatly reduced. All interpolation must take into account both the evolutionary development of organisms and the gross thickness of rock strata. It is unlikely that the broad structure of Table 2 for Paleozoic and later time will be greatly altered as a result of future measurements. [J.L.K.]

Bibliography: D. B. Ericson, W. S. Broecker, J. L. Kulp, and G. Wollin, Late-Pleistocene climates and deep-sea sediments, *Science*, 124:385–389, 1956; A. Holmes, The construction of a geological time scale, *Trans. Geol. Soc. Glasgow*, 21:118–152, 1947; J. L. Kulp, Absolute age determination of sedimentary rocks, *World Petrol. Congr. Proc. 5th Congr., New York*, 1959.

Geology

The study or science of the earth. Geology is one of several related subjects commonly grouped as geoscience. Geologists are concerned primarily with rocks that make up the outer part of the earth. An understanding of these materials involves principles of physics and chemistry; geophysics and geochemistry, now important scientific disciplines in their own right, have become essential allies of geology in exploring the visible and deeper parts of the earth. Study and mapping of surface forms is shared by geology with geodesy. The study of the earth's waters as related to geologic processes is shared by hydrology and oceanography. Paleontology, the study of records left by animals and plants that lived in past ages, is an essential part of geology, though it involves many fundamental aspects of biologic science.

Geology, to contribute its part to geoscience, has developed a number of branches which may be outlined as follows:

I. Physical geology
  A. Mineralogy, study and classification of minerals
  B. Petrology, study of rocks, their physical and chemical properties, and their modes of origin
  C. Weathering and erosion, study of processes that alter exposed bedrock and shape landforms
  D. Sedimentation, origin and deposition of modern sediments
  E. Structural geology, study of geometry of rock masses, with emphasis on crustal deformation
  F. Economic geology, application of geologic knowledge and principles to solution of practical problems
II. Historical geology
  A. Stratigraphy, systematic study of bedded rocks and their relations in time
  B. Paleontology, study of fossils and their locations in sequence of bedded rocks
III. Geologic mapping, graphic representation of bedrock units defined by physical characteristics and geologic age

Nearly all geologic study seeks to determine an order of events, and a major objective is to work out the full history of the earth and its inhabitants.

PHYSICAL GEOLOGY

Mineralogy and petrology. Minerals are the basic units in the composition of most rocks. Although many hundreds of mineral species are known, comparatively few are important in kinds of rocks that are abundant in the visible part of the earth crust. Thus a few related minerals known as feldspars, together with quartz, are the essential ingredients of granite and its near relatives. Limestone, widely distributed on all continents, consists largely of the one mineral calcite. Each mineral has its peculiar atomic structure which under ideal conditions is expressed in crystal form. But when a substance crystallizes in bulk, crowding of grains growing from neighboring centers prevents formation of recognizable crystals. Modern laboratories have effective devices for resolving the mineral content of rock materials; even the ultramicroscopic particles in clays are clearly defined under the [illegible] microscope, which gives images enlarged [illegible]. For further expla- [illegible], commonly applied to the [illegible] earth, is inherited from a speculative concept that our globe was once a molten mass and still is molten below a shell of solid rock. Though the latter view is no longer tenable, evidence from seismology indicates that an outer part of the earth, no more than a few tens of kilometers thick, differs in physical properties from deeper zones, and the convenient term crust is retained for the distinctive outer shell. See EARTH; SEISMOLOGY.

Known rocks are divisible into three groups: igneous rocks, which have solidified from molte[illegible] matter (magma); sedimentary rocks, made of fragments derived from preexisting rocks, of [illegible]al precipitates, or of organic products; a

morphic rocks derived from either igneous or sedimentary rocks under conditions that brought about changes in mineral composition, texture, and internal structure. *See* PETROGRAPHY; PETROLOGY; ROCK.

Igneous rocks. Igneous rocks are formed as either extrusive or intrusive masses, that is, solidified at the earth surface or deep underground. Both kinds range widely in composition; silica, the most abundant ingredient, varies from about 40% to more than 75%. Intrusive bodies (plutons) that formed at various depths are most numerous in mountain zones for two reasons: first, mountain belts have been much deformed, and abundant evidence indicates that crustal disturbance has favored igneous action; second, uplifts in mountain lands have permitted erosion to depths at which plutonic masses were formed. Doubtless some large plutons solidified in reservoirs that formerly supplied volcanoes in action.

Volcanic materials are erupted through two kinds of openings—central vents and long fissures. Central eruptions build up conical mountains; the materials are in large part products of explosion and consist of cinders and ash, with interspersed lava flows. Lavas issuing from fissures have built up vast fields of volcanic rock, chiefly the dark type known as basalt. *See* IGNEOUS ROCKS; MAGMA; PLUTON; VOLCANO.

Sedimentary rocks. Bedrock exposed to air and moisture is broken into pieces, large and small, which are moved by running water and other agents to lower ground, and spread in sheets over river flood plains, lake bottoms, and sea floors. Dissolved matter is carried to seas and other water bodies, and some of it is precipitated chemically and by action of organisms. The material deposited in various ways becomes compacted, and in time much of it is cemented into firm rock. Generally the deposition is not continuous but recurrent, and sheets of sediment representing separate events come to form distinct layers of sedimentary rock. The individual layers are beds or strata, and the rocks are described as stratified.

Large areas in every continent are underlain by sedimentary rocks that represent deposits during many periods of the earth's history. In part these bedded rocks are nearly horizontal, as they were originally (Fig. 1); but in many places, particularly in mountain belts, they have been deformed.

The principal kinds of sedimentary rock are conglomerate, sandstone, siltstone, shale, lime-

Fig. 1. Nearly horizontal sedimentary rocks about 4000 feet thick along Colorado River canyon, Arizona. (Spence Air Photos)

Fig. 2. Effects of erosion and transport by valley glaciers in mountains of southern Alaska. (Alaskan Aerial Survey, U.S. Navy)

stone, and dolomite. Many other kinds, less important quantitatively but with large practical value, include common salt, gypsum, phosphate, iron oxide, and coal. See SEDIMENTARY ROCKS.

Metamorphic rocks. These rocks have been developed from earlier igneous and sedimentary rocks by heat and pressure, most effectively in mountain zones and near large masses of intrusive igneous rock. Thermal metamorphism results from rising temperature, often with addition of new elements by circulating fluids. Common effects are hardening and crystallizing of the affected rock, with changes in mineral composition. Dynamic metamorphism results from shearing stresses in rocks subjected to high pressures; effects are development of cleavage planes and growth of platy (platelike) minerals in parallel arrangement.

The common metamorphic rocks are in the two general classes, foliated (including slate, phyllite, schist, gneiss) and nonfoliated (including marble and quartzite). See METAMORPHIC ROCKS; METAMORPHISM.

Weathering and erosion. Bedrock at and near the earth's surface is subject to mechanical and chemical changes in a complex process called weathering. Blocks and small chips that become detached are especially vulnerable to chemical attack, which makes radical changes in the mineral content. Some soluble products of alteration are removed by percolating water, and in the less soluble residue the most common constituent is clay, which is the basis of soil. The effectiveness of weathering and the nature of its products are controlled by climate, topography, kinds of bedrock, and other variables. Organic processes play a major role in chemical weathering, which is most effective under conditions that favor development of bacteria, plants, and ground-dwelling animals. See WEATHERING PROCESSES.

Weathering prepares the way for removal of rock materials and reshaping of land surfaces by several agents of erosion. The most obvious of these agents is running water, which during a single rainstorm may cut deep gullies into plowed fields and sweep vast quantities of soil, sand, and coarser debris into brooks and eventually into channels of major streams. Abrasion by such moving loads deepens and widens stream channels in hard bedrock. Study of drainage systems brings conviction that even the largest and deepest valleys (Fig. 1) have been fashioned by the action of running water.

Soil on slopes, even those covered with grass and other vegetation, creeps slowly downward. dislodged from cliffs build steep masses rock which slowly migrate downslope. in mountain lands great masses, in

material and bedrock, rush down as landslides. *See* MASS WASTING.

Water moving through underground openings dissolves and carries away great quantities of material. Caverns, large and small, are a conspicuous result. In high latitudes and in some mountain regions glaciers (Fig. 2) are powerful eroding agents. In arid regions quantities of sand and dust are moved by wind with some consequent abrasion of bedrock. Large-scale erosion along the coasts of seas and lakes is performed by waves and currents. *See* CAVE; DESERT EROSION FEATURES; GLACIATED TERRANE; GROUND WATER; SHORE PROCESSES.

Each major agent of erosion fashions characteristic features in landscapes. The net tendency of erosion is to reduce the height of land masses. Various stages in the history of reduction are indicated by forms of valleys and slopes and by relations of land surfaces to the underlying bedrock. Some wide regions have been uplifted and the streams rejuvenated after an advanced stage was reached in a cycle of erosion. *See* GEOMORPHOLOGY.

Sedimentation. An understanding of sedimentary rock is all-important in geology, and sediments now being deposited provide an essential key. On the basis of depositional environment, sediments are assigned to three categories: terrestrial, those laid down on lands; marine, those deposited on sea floors, and mixed terrestrial-marine, those laid down in transitional zones such as deltas, marine estuaries, and areas between high and low tide. In each major group the sediments are further described as clastic (consisting of rock fragments), and chemical (formed either as inorganic precipitates or partly through organic agencies). Study of modern sediments in the several environments takes account of physical peculiarities and also the included remains of organisms.

Terrestrial sediments, widespread and highly varied, are laid down chiefly through the agencies of mass wasting, running water, glacier ice, and wind. The deposits are in large part temporary, as the tendency is for them to shift seaward in continued erosion of the lands. Some subsiding basins under semiarid climates and with no through-flowing streams are retaining all sediments they receive.

A classification of marine sediments according to their source follows:

1. Derived from lands and contributed by (*a*) streams, (*b*) wave erosion of coasts, (*c*) winds, (*d*) floating ice.

2. Formed in the sea by (*a*) shells and skeletons of marine animals and plants, (*b*) chemical precipitation.

3. Fragmental material erupted from volcanoes.

4. Particles of meteorites from outside the earth. *See* MARINE SEDIMENTS.

Structural geology. Geometric study of rocks distinguishes primary structures, acquired in the genesis of a rock mass, and secondary structures that result from later deformation. Significant features in sedimentary rocks make them especially valuable for registering later changes in form.

Within historic time abrupt movements have occurred along large breaks (faults), with instantaneous displacements as much as tens of feet (Fig. 3). Such movements are attended by strong earthquakes. Another type of movement, now in progress but at an extremely slow rate, is doming up the surface, as in Scandinavia, on a regional scale. Similar ancient warping is evident in the gentle bending of sedimentary strata, although the land surfaces that were deformed have been destroyed by erosion. In addition to broad warps, the principal kinds of structural features that record deformation are folds, joints, faults, cleavage, and unconformities. *See* STRUCTURAL GEOLOGY.

Evidence of the most pronounced crustal deformation is found in mountain belts, where erosion has exposed exceptionally thick sections of sedimentary rocks. These rocks record long histories of slow subsidence and sedimentation, interrupted by large-scale deformation and uplift. *See* DIASTROPHISM; OROGENY.

Major relief features of the earth reflect differences in density of the underlying rocks. Continental rocks, diverse but including great masses with granitic composition, have appreciably lower average density than the basaltic rocks of ocean floors. Great mountain blocks such as the Alps and the Himalayan chain represent thickened parts of continental masses. The condition of approximate balance among diverse parts of the crust is called isostasy (equal standing). *See* ISOSTASY.

Economic geology. A general knowledge of geology has many practical applications, and large numbers of geologists receive special training for service in solving problems met in the mining of metals and nonmetals, in discovering and producing petroleum and natural gas, and in engineering projects of many kinds. *See* ENGINEERING GEOLOGY; PETROLEUM GEOLOGY.

HISTORICAL GEOLOGY

The legible history of the earth is read chiefly from sedimentary rocks, which record a sequence of events, changing physical environments, developments in plant and animal life, effects of crustal movements. Additional records are supplied by volcanic rocks, which in many areas are interlayered with and grade into sedimentary strata; by relations of intrusive igneous bodies to older and younger rocks; and by erosion surfaces, some displayed in present landscapes, others revealed in exposures of unconformities. The long history includes r [illegible] changes i [illegible] ical geography, featurin [illegible] s cor [illegible] land and sea; the [illegible] and [illegible] f successive mou [illegible] ; a [illegible] of living form [illegible] on l [illegible] of events t [illegible] e is [illegible] strath [illegible] in [illegible] g

Fig 3. Fault near Great Bear Lake, Canada. This fault zone, etched by erosion, is traceable for 80 miles. (Royal Canadian Air Force)

Stratigraphy is the systematic study of stratified rocks. Paleontology is the study of fossilized plants and animals, with regard to their distribution in time. These two complementary disciplines provide much of the basis for geologic history.

Stratigraphy. Studies of consolidated sedimentary rocks in comparison with the many kinds of modern sediments provide a basis for recognizing conditions under which the older deposits were formed—whether on land, on sea floors, or in transitional zones such as deltas or lagoons. Generally a much closer interpretation is possible; bodies of rock are confidently classified as deposits on flood plains, at margins of glaciers, in large lakes, in shallow seas near shore, or in deep-sea troughs. Each distinctive type of deposit presents a facies. A kind of rock that is essentially uniform over a considerable area constitutes a lithofacies. An assemblage of fossils that is nearly uniform in a large unit of sedimentary deposits, indicating an environment suited to certain forms of life, marks this unit as a biofacies. Deposits formed at the same time may differ greatly both in lithofacies and in biofacies, reflecting differences in topography, climate, and other items of environment.

A fundamental principle in stratigraphic studies, known as the law of superposition, states that in a normal sequence of strata any layer is older than the layer next above it (see Fig. 1). This elementary law is of fundamental importance in a study of many mountain zones where thick sections of strata have been overturned, even completely inverted, and can be resolved only through criteria that indicate original tops of beds. Close matching of the many kinds of modern sediments with materials in sedimentary rocks formed over an immense span of time has established the uniformitarian principle, which holds that processes now operating on the earth have operated in fairly uniform fashion through the ages. See STRATIGRAPHY.

Each continent (other than ice-buried Antarctica) has at least one wide lowland or platform that has been occupied repeatedly by seas and is now mantled with little-deformed marine strata. The total deposit on each platform represents a long span of time but ranges from only a few hundred to a few thousand feet thick. Along a margin of each platform the section of strata is much thicker, more varied in character, and strongly deformed. This belt of thick deposits records long-continued subsidence of a geosyncline; the strata then were folded and faulted, and elevation of the belt formed high mountains. Examples are the Appalachian Mountains, the younger Rocky Mountains, and the still younger Alpine-Himalayan chains. See GEOSYNCLINE; TECTONIC PATTERNS.

Paleontology and the scale of time. At any one locality a sequence of sedimentary beds, from older to younger, can be determined through physical evidence. Persistence of some peculiar units may establish approximate correlations through moderate distances, occasionally hundreds of miles. But the accepted key to relative dating of stratigraphic units and to confident world-wide correlations is supplied by fossils of animals and plants, which record progressive evolution in living forms from ancient to recent times. Many of the oldest known sedimentary rocks are now highly metamorphosed, and any fossils they may have held must have been obliterated. Some thick sections of very old strata that are not appreciably altered have yielded only sparse indications of life, such as patterns of marine algae and burrows made by worms or other lowly forms. Successively younger groups of strata

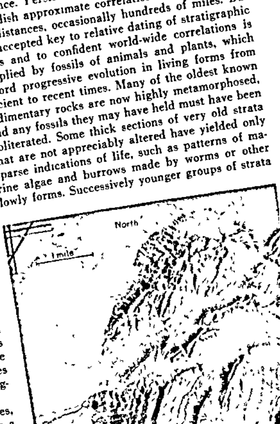


Fig. 4. Vertical air view of mountains east of Vegas, Nevada. (USGS)

hold abundant fossils of marine invertebrates, marine fishes, land plants that progressed from primitive to more modern kinds, reptiles, birds, small mammals, followed by diverse and generally more advanced kinds of mammals, primitive men, and finally modern man Some forms evolved rapidly, and short-lived species that were equipped to become widely dispersed are of greatest value for correlation. *See* PALEOBOTANY; PALEONTOLOGY.

A geologic scale of time (see table) is based partly on physical evidence but more largely on the paleontologic record as is suggested by names of the major divisions (Paleozoic, ancient life; Mesozoic, medieval life; Cenozoic, recent life). Absolute dates are determined by analyses of critical isotopes, collected at strategic locations. This dating is known as geochronometry. *See* GEOCHRONOMETRY; GEOLOGICAL TIME SCALE; see also articles on the divisions of geologic time named in the table.

GEOLOGIC MAPPING

A geologic map represents the lithology and, so far as possible, the geologic age of every important geologic unit in a given area. Each distinctive unit that can be shown effectively to the scale of the map is a geologic formation. A good topographic

Geologic column and scale of time

Periods of time Systems of rocks	Epochs of time Series of rocks	Distinctive records of life	Absolute dates (in years before present)
		CENOZOIC ERA	
Quaternary	Recent	Modern man*	11,000
	Pleistocene	Early man	600,000
Tertiary	Pliocene	Large carnivores	12,000,000
	Miocene	Whales, apes, grazing animals	20,000,000
	Oligocene	Large browsing animals	35,000,000
	Eocene	Rise of modern floras	55,000,000
	Paleocene	First placental mammals	65,000,000
		MESOZOIC ERA	
Cretaceous		Last of dinosaurs	90,000,000
		Last of ammonites	120,000,000
		Rise of flowering plants	140,000,000
Jurassic		Toothed birds	155,000,000
		Flying reptiles	170,000,000
		First primitive mammals	185,000,000
Triassic		Rise of dinosaurs	200,000,000
		Rise of ammonites	215,000,000
		Rise of cycads	230,000,000
		PALEOZOIC ERA	
Permian		Primitive reptiles	245,000,000
		Last of trilobites	260,000,000
		Glossopteris flora	275,000,000
Carboniferous periods		Spread of amphibians	
Upper (Pennsylvanian)		Great coal forests	310,000,000
		Climax of spore-bearing plants	
Lower (Mississippian)		Abundant sharks	350,000,000
		Climax of crinoids and blastoids	
Devonian		First forests	365,000,000
		Rise of ferns	385,000,000
		Earliest known amphibians	400,000,000
Silurian		Appearance of land plants	
		First known scorpions	(420,000,000)
Ordovician		Expansion of brachiopods and corals	
		Appearance of primitive fishes	440,000,000
		Climax of trilobites	460,000,000
Cambrian		Rise of cephalopods	490,000,000
		Abundant trilobites	510,000,000
		Many kinds of shelled invertebrates	540,000,000
		PRECAMBRIAN TIME†	
No basis for worldwide subdivisions		Marine algae, worm burrows, other simple forms	1,000,000,000
			1,400,000,000
		Abundant carbon of organic origin	1,800,000,000
			2,600,000,000
			3,400,000,000
			4,500,000,000

* Recent and Late Pleistocene.

† Figures for Precambrian time based on well-dated orogenic events that occurred over large geographic areas.

base map is essential for representing relations of bedrock to forms of the land surface. Cooperation of students with specialized qualifications—for example in petrology, paleontology, structural geology—is required for accurate mapping and description of complex areas. Large organizations, such as Federal geological surveys and some commercial companies, have diversified personnel, laboratories equipped for varied analyses, and special field equipment. Aerial photographic surveys serve as a guide in field work and help in plotting accurate locations on maps. Photogeology is a technique that uses photographs (Fig. 4) for constructing preliminary geologic maps, which are checked and corrected by geologists working on the ground.

A completed geologic map should indicate important structural details such as inclinations of strata, locations of faults, and axial traces of folds. Usually the map is supplemented by vertical sections on which structural features seen at the surface are projected to limited depths. Maps of small scale may represent sedimentary rocks only according to the systems to which they belong. With larger scale a given system may be represented by several formations, each recording an important episode in the history of the region. In some European countries geologic mapping has been completed to fairly large scales; but in all continents great areas are still unexplored geologically or have been mapped only in reconnaissance fashion. [C.R.L.]

Bibliography: E. B. Branson et al., *Introduction to Geology*, 3d ed., 1952; T. H. Clark and C. W. Stearn, *The Geological Evolution of North America*, 1960; E. C. Dapples, *Basic Geology for Science and Engineering*, 1959; C. O. Dunbar, *Historical Geology*, 2d ed., 1959; W. H. Emmons et al., *Geology, Principles and Processes*, 5th ed., 1960; R. M. Garrels, *A Textbook of Geology*, 1951; J. Gilluly et al., *Principles of Geology*, 2d ed., 1959; J. A. Jacobs, R. D. Russell, and J. Tuzo Wilson, *Physics and Geology*, 1959; L. D. Leet and S. Judson, *Physical Geology*, 2d ed., 1958; C. R. [illegible]

and K. E. Caster, *Geology*, 1952.

Geomagnetic storm

The name given to a large disturbance of the earth's magnetic field. Magnetic storms are of world-wide extent. In many cases they begin suddenly all over the earth, simultaneously within about a minute. The field variations at the earth's surface during a storm are greatest near the auroral zones, north and south. Great magnetic storms are accompanied by outstanding auroral displays, which are then seen in lower latitudes than usual. Magnetic storms go through a fairly regular "life cycle," although irregular local variations are superposed on the more regular features. Between the auroral zones, the main regular change affects chiefly the horizontal magnetic intensity H. This first increases for a few hours, by 10 γ or more ($\gamma = 10^{-5}$ gauss), then decreases to a minimum—up to 100 γ or even 200 γ below normal. This minimum is attained most rapidly (12–15 hours) in the greatest storms; in weak storms the minimum is very flat, and is attained in 1–2 days. Then H returns to normal, during several days.

In polar regions, the intense variations especially affect H and Z. The Z, or vertical, variations are opposite on opposite sides of the zones. The polar variations seem to be produced by large electric currents in the ionosphere. Concentrated currents (electrojets) flow partly eastward and partly westward, along the auroral zones. The westerly part is the more intense, and may be of the order of 10^6 amperes. The currents complete their path mainly over the polar caps enclosed by the zones.

The field-changes in the belt of the earth between the zones may be produced partly by electric currents far above the earth, at a height of a few earth radii. It seems likely that the main phase of a storm, that is, the decrease of H, is caused by currents in the Van Allen belts, which were discovered and explored by moon rockets and satellites launched during the International Geophysical Year. Magnetic storms seem to be caused by streams or clouds of ionized gas ejected from the sun, with a speed that can carry them to the earth in about a day. These streams or clouds probably also "feed" the Van Allen belts, and produce auroras. There are many important time relationships between magnetic storms and intrinsic changes on the sun, especially solar flares. *See* GEOMAGNETISM; SUN. [S.C.]

Bibliography: *See* GEOMAGNETISM.

Geomagnetic variations, transient

Nonperiodic, irregular, and generally short-term variations in elements of the earth's magnetic field, particularly magnetic storms or geomagnetic storms occurring most commonly at times of unusual sunspot activity (*see* SUNSPOT).

Transient geomagnetic variations are recorded continuously at about 150 magnetic observatories widely distributed over the globe. The magnetic elements usually recorded are declination or direction D, horizontal intensity H, and vertical component Z (*see* GEOMAGNETISM). In most months there are several days, at least, when the daily curves show a regular variation. This is called the solar daily geomagnetic (gm) variation, and is commonly denoted by S_q (S for solar, q for quiet). It is the local manifestation of a changing (S_q) field superposed on the main magnetic field. The S_q field is produced mainly by electric currents in the E layer of the ionosphere. Frequently, during a solar flare, additional ionization and electrical conductivity are produced in the D layer, immediately below the E layer, over the sunlit hemisphere (*see* AERONOMY; ATMOSPHERE; IONOSPHERE). Then additional electric currents, of similar type, flow in the D layer. They briefly magnify the S_q field over the sunlit hemisphere

causing a "bay," denoted by S_{qa} (a for augmentation) in the magnetic records. The electric currents in the E layer are induced by dynamo action of periodic air motion in the ionosphere, across the lines of force of the main geomagnetic field.

Lunar tidal air motions in the ionosphere add to this dynamo action, and produce a weak additional system of electric currents, and an additional weak varying magnetic field. Its changes at the earth's [illegible]

to the field and electric current system of which they are the manifestation. The L currents and field and recorded variations are more intense during the sunlit hours, because the ionosphere is more strongly ionized and more electrically conducting over the sunlit than over the dark hemisphere of the earth. At most observatories, the presence of the L variation cannot be detected by inspection of the magnetic records.

The daily range of the S variation is of the order of 40 γ ($\gamma = 10^{-5}$ gauss); the range of the L variation is smaller by a factor of 10 to 15.

At Huancayo Observatory, in Peru, almost on the magnetic equator, in the horizontal intensity H the S_q variation is abnormally large (daily range about 100 γ), and the abnormality in L is even larger. There the presence of the L variation in H is readily detectable in the records, since L is about one-third as large as S_q.

At times, the regular daily magnetic variation is disturbed. When intense, such disturbance is called a magnetic storm (see GEOMAGNETIC STORM).

The S_q and L fields, and likewise the field of magnetic disturbance produced above the earth's surface, induce electric currents within the earth. The magnetic field of these currents contributes to the observed S, L, and disturbance variations. Analysis of these variations serves to separate their parts of external and of internal origin, and supplies information about the electrical conductivity within the earth to depths of several hundred kilometers.

The S field is greater in years of sunspot maximum than in those of sunspot minimum. [S.C.]

Bibliography: See GEOMAGNETISM.

Geomagnetism

A term signifying both the magnetism of the earth and the branch of science that deals with the earth's magnetism. The term geomagnetism is now given some preference over the older and longer term terrestrial magnetism.

Main geomagnetic field. This is specified at any point O by its vector magnetic intensity, **F**. Its direction is that of the line $P'OP$ from the negative end P' to the positive end P of a magnetized needle $P'P$, perfectly balanced and freely pivoted about O, when in equilibrium. The positive pole P' is the one that at most places on the earth takes the more northerly position. Over most of the earth's Northern Hemisphere, P will be below O; the needle is said to dip below the horizontal by an angle I, called the magnetic inclination.

Over about half the earth, however, P will be above O; the inclination I (or magnetic dip) is then reckoned as negative. This is the case over most of the Southern Hemisphere. The value of I thus ranges from 90° to −90°. A point where $I = \pm 90°$ is called a magnetic pole of the earth.

Magnetic poles and equator. There are two main magnetic poles; their approximate positions in 1955 were 73.5°N, 100°W and 71.5°S, 151°E. In a few places on the earth, near strongly magnetized mineral deposits, there may be local magnetic poles.

The distribution of I over the earth's surface can be indicated on a globe, or on a map (with any kind of projection), by lines called isoclinic lines, or isoclines along each of which I has the same value. The isocline for which $I = 0$ (where the balanced magnetized needle rests horizontal) is called the magnetic equator. Figure 1 shows isoclines over a large part of the earth for the epoch 1945.

Magnetic declination. A compass needle is a magnetized needle weighted at one end so as to be horizontal when freely pivoted about its center. When free to turn in the horizontal plane, its direction relative to geographic north is called the magnetic declination D. This is reckoned positive to the east and negative to the west; alternatively, declinations can be specified by a positive (eastward) value ranging up to 360°.

Over the greater part of the earth D is numerically less than 90°, but along any small circuit around the magnetic poles it takes all values. At the magnetic poles themselves the compass needle takes no definite direction. At the geographic (gg) poles the needle takes a definite direction, but as the northward direction changes through a whole revolution along any small circuit around the pole, D likewise takes all values around the circuit.

The distribution of the declination over the earth's surface can be indicated by lines along each of which D is constant. These are called isogonic lines, or isogones. For the reasons just stated, the two magnetic poles and the two geographical poles are points toward which isogones converge, for all values of D. Figure 2 shows isogones for a large part of the earth, for the epoch 1945. The isogones for which $D = 0$ are called agonic lines. Along these lines, the compass points to true north.

The distribution of D over the earth can also be indicated by magnetic meridians, which at each point P have the direction of the compass needle at P. These lines extend from the south magnetic pole to the north, and have no complication (as do the isogones) at the geographic poles. They indicate, more clearly than the isogones, the general distribution of the compass direction over the earth. But the isogonic map is more convenient in enabling the compass direction at any point to be read or estimated without using an angle measurer. Hence it is the one used by navigators and travellers on sea and land, and in the air.

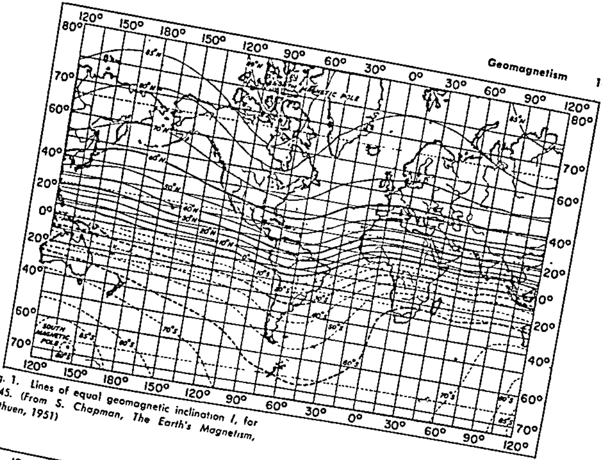


Fig. 1. Lines of equal geomagnetic inclination *I*, for 1945. (From S. Chapman, *The Earth's Magnetism*, Methuen, 1951)

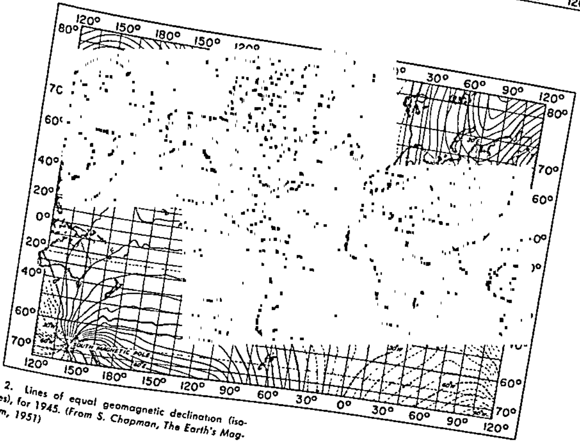


Fig 2. Lines of equal geomagnetic declination (isogones), for 1945. (From S. Chapman, *The Earth's Magnetism*, 1951)

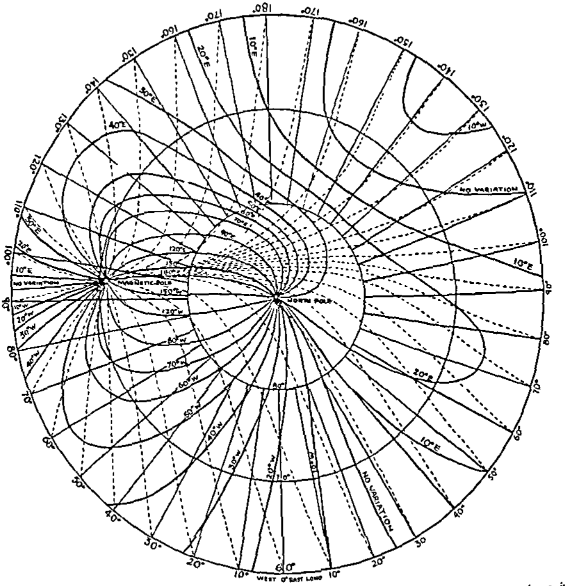


Fig. 3. Lines of equal declination (full lines) and magnetic meridians (broken lines), for 1922, for the Arctic region north of 60° latitude. The agonic lines are marked "No Variation." *(After H. Spencer Jones, in S. Chapman and J. Bartels, Geomagnetism, 2 vols., Oxford, 1940)*

Figure 3, which shows both isogones and magnetic meridians in the region around the north geographic and magnetic poles, illustrates this contrast between them.

Intensity patterns. The strength F of the vector intensity **F** is the third quantity which, together with I and D, completely specifies **F**. It is expressed in a unit which in geomagnetic literature is called the gauss, after the German mathematician, astronomer and physicist C. F. Gauss (1777–1855) who first showed how F could be measured in units of length, mass, and time. This was the first nonmechanical entity so to be measured; Gauss' innovation was a landmark in the history of physical measurement. Physicists and electrical engineers distinguish the gauss from the oersted, the unit by which they would express the magnetic force. The numerical difference between the two, in geomagnetism, is negligible.

The symbol for the gauss is Γ; much use is made in geomagnetism of a unit called the gamma (symbol γ), which is 10^{-5} Γ.

The distribution of the magnetic intensity F over the earth can be indicated, as for I and D, by lines of constant F. They are called isodynamic lines. They are shown for the epoch 1915 in Fig. 4. The value of F in general increases with increasing latitude north or south; its distribution, however, is not simple. In low latitudes, it has a minimum of about 0.25 Γ off the east coast of South America; in the tropics near Indonesia it is over 0.35 Γ near Indonesia. Near the north magnetic pole, its value rises to 0.7 Γ.

The vector intensity **F** can also be specified by its vertical component Z and its horizontal vector component **H**; the latter can be specified by its direction (given by D) and the horizontal intensity H or, alternatively, by its north and east components X, Y.

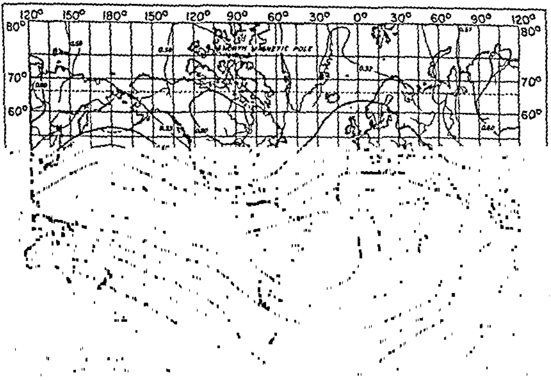


Fig. 4. Lines of equal geomagnetic intensity, *F*, for 1945. (From S. Chapman, *The Earth's Magnetism*, Methuen, 1951)

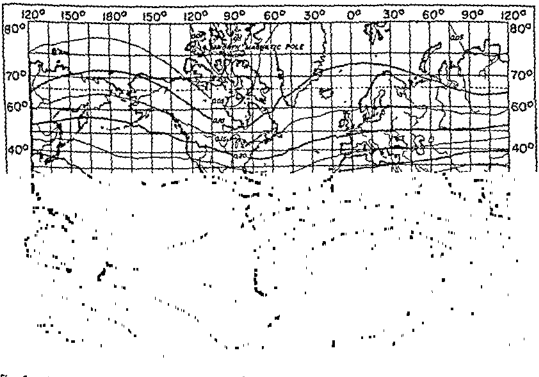


Fig. 5. Lines of equal horizontal geomagnetic force for 1945. (From S. Chapman, *The Earth's Magnetism*, Methuen, 1951)

Clearly $Z = F \sin I, H = F \cos I, X = H \cos D$, and $Y = H \sin D$; Z, like I, is thus reckoned positive when downward, and Y, like D, is positive when eastward.

Geomagnetic elements. All the magnitudes F, H, X, Y, Z, I, and D are together called geomagnetic elements. Lines on a map along which any element has a constant value are called isomagnetic lines, and the map is called an isomagnetic chart. Figure 5 shows an isomagnetic chart for H for the epoch 1945.

Isomagnetic maps for the whole earth, or for any large part of it, cannot show the finer details of the field distribution. Maps for smaller areas, such as for one of the American states or of a European country, can show such details, if the areas have been closely surveyed magnetically. In some parts of such maps, the course of the lines may depart considerably from the smooth spacing indicated on world maps. They can show regions of local magnetic anomaly, where the geomagnetic field is disturbed by large

Kursk region of Russia. The strips are 60 km apart, and run in parallel from NE to SW. The most disturbed part of the major (northerly) strip is only 2 km wide, although the strip is 250 km long; Z is everywhere above normal, and ranges up to 1.9 Γ.

Mathematical probe of spherical harmonics. The earth's surface magnetic field can be expressed in mathematical terms by applying a process called spherical harmonic analysis. The data used are magnetic survey measurements of the magnetic elements at a sufficient number of points on the earth's surface. The analysis provides separate expressions for the parts of the surface field that originate respectively within the earth and above it. It proves that the main field comes almost entirely from within. Any part that is of external origin is too small to be determined accurately in this way from the available data; its strength is probably less than 300 γ at the earth's surface, and perhaps only 20–30 γ.

The main term in the spherical harmonic expression corresponds to the field of a uniformly magnetized sphere, or the equivalent at outside points of a point dipole at the earth's center. Hence this part of the geomagnetic field is called the dipole field. The intensity of magnetization for a sphere the size of the earth, to give this dipole field, would be about 0.08 Γ. The field could also be produced by a surface distribution of electric current around the earth, of total amount 1.5×10^9 amperes, distributed proportionately to the sine of the latitude. The moment of the equivalent dipole is 8.1×10^{25} Γ-cm³. These three possible systems that could produce the dipole field all differ from the actual source of the field.

The earth's diameter along the direction of the dipole (which is also the direction of magnetization for the alternative model of a uniformly magnetized sphere), is called the geomagnetic (gm) axis. Its ends on the earth's surface are called the gm or axis poles. The polarity of the northern pole is negative; that of the southern is positive. The northern gm pole is conveniently marked B on maps, and the southern, A; these letters may stand for boreal and austral. The gg (geographic) position of B is 78.5°N, 69°W, and of A, 78.5°S, 111°E. The obliquity of the gm to the gg axis is therefore 11.5°. The gm pole B is 1280 km from the N gg pole and 1160 km from the north magnetic pole; the corresponding southern distances are 1280 km and 1350 km. The geomagnetic axis has shown no change of direction large enough to be estimated within a degree or two since it was first determined about a century ago. The position of any point P may be expressed in geomagnetic coordinates relative to the gm axis; the gm colatitude is the angle BOP (O being the earth's center); the gm longitude is measured eastward from the part of the common gg and gm meridian through B, that lies south of B. The geomagnetic equator, from which gm latitude is measured north toward B or south toward A, is in the diametral plane perpendicular to the gm axis.

A closer approximation to the earth's surface field is the field of a point dipole not at the earth's center. The eccentric dipole must have the same direction and magnetic moment as the centered dipole but is displaced from O by about 340 km, toward a point in the Pacific Ocean northeast of New Guinea, at 6.5°N and 161.8°E. The eccentric dipole field is only a moderate improvement on the centered dipole field, as an approximation to the actual surface field; there still remain regional differences in the horizontal component, amounting in several places to over 0.1 Γ.

The origin of the geomagnetic field is discussed later, in connection with the second main branch of geomagnetic science.

Secular magnetic variation. This is a slow change of the earth's field. In any magnetic element at a particular place, the variation may be an increase or a decrease; its rate is unpredictable, and not constant in either magnitude or sign. It necessitates continual redrawing of the isomagnetic maps. The distribution of its own rate, for any element, can be indicated on maps called isoporic, by lines (isopors) along which the rate is constant.

The pattern of such isopors is more complex than that of the isomagnetic lines for the same element. There may be several regions of maximum or minimum change, producing numerous oval systems of isopors (see Fig. 6, giving Z isopors based on an interval of 38 years). Whereas the main field is a planetary property, the secular variation is regional. Moreover isoporic maps change much more than isomagnetic maps from one decade to the next. Regions of maximum change move, disappear, or newly appear. An example of long-continued changes of the field direction is given in Fig. 7, for London. From 1576 to about 1820 the

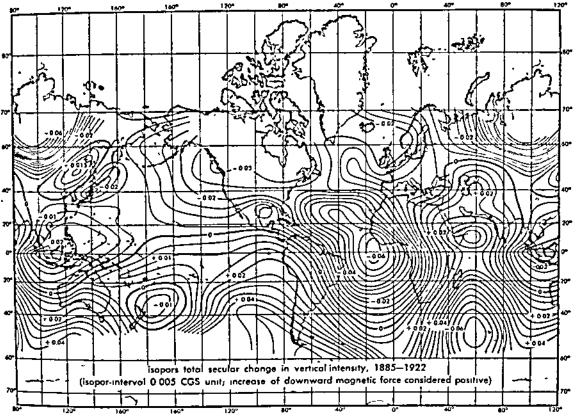


Fig. 6. Lines of equal rate of change (isopors) of the vertical geomagnetic force, *Z*, over the interval from 1885 to 1922. The isopors are drawn at 500-γ intervals.

(*After A. G. McNish, in S. Chapman and J. Bartels, Geomagnetism, 2 vols., Oxford, 1940*)

compass there turned from 11°E to 24°W, while the dip first increased from 71° by 3°, until 1700; since 1700 until now it has decreased again, to about 67°. Since 1800 the compass has swung eastward through about 16°. Figure 7 shows these changes, along with some directions and trends for earlier centuries (indicated by numbers, that is, 14 indicates fourteenth century), estimated from the magnetization of bricks in kilns whose period can be dated. This is an example of paleomagnetic research, which aims to extend knowledge of the past geomagnetic changes by centuries, millennia, and more.

Figure 7 suggests a possible cycle of change of direction at London, with a period of 4–5 centuries. But the changes elsewhere are inconsistent with any regular period of secular geomagnetic change.

Another example of a long-continued secular trend is the movement of the agonic line, which in Fig. 2 crosses the Equator at about 65°W. The course of this agonic line has been followed longer than that of the more complicated looped agonic line that traverses Asia. In 1550 it crossed the Equator at about 20°E, and since then it has moved continuously westward through about 90° in about four centuries. This is one token of a noticeable tendency—the westward drift—of isomagnetic and isoporic features. Another large-scale feature of the secular magnetic variation is a decrease of the earth's magnetic moment by about 4% during the last century. Present evidence suggests that this decrease has now ceased and that the moment is beginning to increase.

Geomagnetic measurements. To keep our knowledge of the earth's magnetic field abreast of its secular changes, the world must be magnetically surveyed from time to time. In the past this has generally been done, on land and sea, by measuring *D*, *H*, and *I*. To measure *D*, it is necessary to determine true north by astronomical observation; for *I*, the horizontal direction must be determined by a good level. Then a weighted compass needle (declinometer) or a dip circle with balanced needle will indicate *D* and *I*; for real accuracy, great care and suitable reversals are required. To measure *H*, the original method devised by Gauss was by means of a magnet, whose oscillations and power to deflect another magnet were observed. The instrument used is called a magnetometer. *See* COMPASS, MAGNETIC; DECLINOMETER; MAGNETOMETER.

Another form of inclinometer to measure *I* which has almost entirely displaced the dip circle has a rotating coil whose axis of rotation is turned until rotation in the earth's field produces no in-

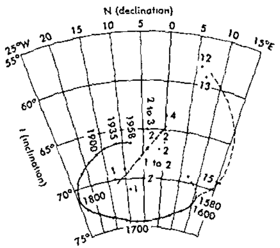


Fig. 7. The geomagnetic direction at London, as observed from 1580 to 1958; the magnetic declination and the dip are shown, mostly at 20-year intervals. In addition, points are shown indicating the direction for earlier centuries, inferred from studies of magnetized bricks in old Roman-British and later kilns. The broken lines are conjectural; the straight broken line shows the estimated changes of direction in Roman times. (*After L. A. Bauer, R. M. Cook, and J. C. Belshe*)

eter, and earth inductor compass. *See* EARTH INDUCTOR

Geomagnetic measurements in field surveys aim at a lower accuracy than those in fixed magnetic observatories, where the transient field changes are continuously recorded (*see* GEOMAGNETIC VARIATIONS, TRANSIENT). Measurements at the best observatories are absolute in the sense that they use instruments directly calibrated in terms of the units of length, mass, and time. Where, as in field surveys, the instruments used have been calibrated only by comparison with such absolute instruments, the observations are relative.

At some of the best magnetic observatories, absolute magnetometers are used; these are based on the comparison of the geomagnetic field with the field of a known electrical current in an accurately measured coil.

Other types of magnetometer depending partly on electrical currents are the flux gate magnetometer and the proton magnetometer.

In some countries arrangements are made for repetition of the magnetic survey at suitable intervals, such as every 15–30 years; other countries lag in this respect. Local anomalies are found to persist, superposed on the changing regional magnetic field.

Recent concepts and developments. Magnetic surveys can now be made by air. A new world magnetic survey is planned as a deferred but integral part of the International Geophysical Year. Given feasible technical developments, such a survey could be made quickly by satellite-borne instruments, if the orbit passed over or near the earth's poles. *See* AEROMAGNETIC SURVEYING.

The source of the main geomagnetic field and of its secular variation is within the earth, according to presently accepted theories. Although slow by some standards, the secular variation is very rapid by geologic standards. It cannot reasonably be ascribed to changes in or not far below the earth's crust. It now seems probable that the main field is caused by electric currents in the earth's liquid core, below about 2900 km (1800 miles). By slow convective movements, electric currents are produced in the core; these maintain the magnetic field, as in a self-exciting dynamo. Rather large-scale eddies in the convective motion produce the regional features of the main field, and their changes produce the secular magnetic variation. These changes have no apparent relation to the broad features of geography and geology.

According to this theory, the geomagnetic field will remain of mainly dipole character nearly down to the surface of the core, with the lines of magnetic force lying nearly in planes through the geomagnetic axis. The magnetic intensity will increase inversely as the cube of the distance from the center, to nearly 5 Γ at the axis poles of the core; but the regional irregularities will become more prominent as the core is approached.

Inside the core the lines of force are probably much twisted around the earth's axis, and the intensity may approach 500 Γ. There will be a reaction between the magnetic field and the electrical currents in the core, tending to produce a westward drift of the core (and hence of the secular magnetic variations) relative to the outer solid part of the earth.

The electrical conductivity of the core required by these theories is of order 3×10^{-6} emu, or 3×10^{6} ohm-cm. This is far greater than that of the upper layers of the earth; in ohm-cm, for the oceans the conductivity is 40–60; for dry rock, of order 10^{-3}–10^{-4}; at about 600 km depth, it has increased to 10, and it may reach a maximum of about 3000 at 1000 km depth, thence declining to about 10 at the core surface. *See* ELECTRICAL UNITS. [S.C.]

Bibliography: S. Chapman, *The Earth's Magnetism*, 2d ed., 1951; S. Chapman and J. Bartels, *Geomagnetism, 2 vols.*, 1940; J. A. Fleming (ed.), *Terrestrial Magnetism and Electricity*, reprint, 1949; E. H. Vestine et al., *Description of the Earth's Main Magnetic Field and Its Secular Change, 1905–1945*, Carnegie Inst. Wash. Publ. 578, 1947; E. H. Vestine et al., *The Geomagnetic Field*, Carnegie Inst. Wash. Publ. 580, 1947.

Geometry, differential

A branch of mathematics that deals with the intrinsic properties of curves and surfaces in the three-dimensional euclidean space. The intrinsic properties are those which are independent of the geometrical object's orientation or location in space. The subject is also concerned with nets of curves and families of surfaces, these having wide application in the arts.

The space is referred to a rectangular cartesian coordinate system (x,y,z). A space curve may be defined by a pair of independent equations $f(x,y,z) = 0$ and $g(x,y,z) = 0$, or, more meaningfully, by parametric equations (1): $x = x(t)$, $y = y(t)$, $z = z(t)$. In this case the arc length between two points t_0 and t is given by

$$s = \int_{t_0}^{t} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt \qquad (1)$$

Obviously if s is chosen as parameter

$$\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 + \left(\frac{dz}{ds}\right)^2 = 1$$

In this case dx/ds, dy/ds, dz/ds at a point P are the direction cosines of the tangent to the curve at P.

Plane curves. Consider three nearby points P, P_1, P_2. Through them, in general, one plane may be constructed. The limiting position of this plane as P_1 and P_2 approach P is the osculating plane of the curve at P. The limiting position of the circle through P, P_1, P_2 as P_1 and P_2 approach P is the osculating circle. Its center is the center of curvature and its radius is the radius of curvature ρ. The reciprocal of the radius of curvature, κ, is the curvature; its value is

$$\sqrt{\left(\frac{d^2x}{ds^2}\right)^2 + \left(\frac{d^2y}{ds^2}\right)^2 + \left(\frac{d^2z}{ds^2}\right)^2}$$

The perpendicular to the tangent at P in the osculating plane is the principal normal and the perpendicular at P to the osculating plane is the binormal. Another way of defining curvature is this: Let P' be a point near P, arc $PP' = \Delta s$. Let $\Delta\theta$ be the angle between the tangents at P and P'. Then

$$\frac{1}{\rho} = \lim_{P' \to P} \frac{\Delta\theta}{\Delta s}$$

Similarly, if $\Delta\phi$ is the angle between the binormals at P and P',

$$\lim_{P' \to P} \frac{\Delta\phi}{\Delta s}$$

is the torsion $1/\tau$ of the curve at P, where τ is the radius of torsion. The torsion at P is the rate at which the curve is twisted out of the osculating plane so that the torsion of a plane curve is zero. The values of the curvature and torsion in terms of the arc determine the curve up to its position in space. The determination of the parametric equations of the curve, as well as of the solution of most problems dealing with space curves, is accomplished by means of the Frenet equations. These equations give the values of the derivatives of the direction cosines of the tangent, normal, and binormal in terms of these cosines and the curvature and torsion of the curve.

Plane curves are of course special cases of space curves ($\tau = 0$). One of the problems
tensively in classical differential geometry

of nets of curves. Such a net is given by two independent equations

$$x = x(u,v) \qquad y = y(u,v) \qquad (2)$$

For fixed values of v, Eqs. (2) define a family of curves, the u lines, whereas fixed values of u give the v lines. Through each point there is one u line and one v line so that (u,v) may be regarded as curvilinear coordinates in the plane. If this system is to be orthogonal it is necessary that

$$\frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} = 0$$

Equations (2) may also be interpreted as the mapping of the uv plane on the xy plane. In applications one seeks a mapping under which certain properties of figures are preserved. If arc length is to be preserved the mapping must be a motion; if area is to be preserved it is necessary that

$$\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v} = 1$$

Probably the most important mappings are the conformal ones. Under them the angles between lines are preserved; such a mapping is given by $x + iy = f(u + iv)$, where f is any analytic function of the complex variable $u + iv$ ($i = \sqrt{-1}$).

Surfaces. A surface in three-dimensional euclidean space may be given by $f(x,y,z) = 0$ or, more conveniently, by

$$x = x(u,v) \qquad y = y(u,v) \qquad z = z(u,v) \qquad (3)$$

For a fixed value of v, Eqs. (3) describe a curve in the surface, a u line, and similarly for a fixed value of u. Thus (u,v) are curvilinear coordinates of a point in the surface. The most important quantity in the study of surfaces is the arc length of a curve. This is given by

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2 \qquad (4)$$

where E, F, G are functions of partial derivatives of (x,y,z). In the real domain E, G and $EG - F^2$ are nonnegative and may be zero only at singular points of the surface, or at points where the matrix of the partial derivatives of (x,y,z) is of rank less than two. The right-hand side of Eq. (4) is called the first fundamental form of the surface, where E, F, G are the first fundamental quantities of the surface. All applicable surfaces have the same first fundamental form. Thus any cylinder or cone which is developable (applicable to a plane) has the fundamental form $du^2 + dv^2$, where (u,v) are rectangular cartesian coordinates. Under an arbitrary transformation of the surface coordinates (u,v) the fundamental quantities E, F, G transform linearly so that many problems in the study of surfaces reduce to the question whether a
nate system exists on the surface in wh
satisfy desired conditions.

In order to distinguish between
having the same first fundamental
siders the plane tangent to the .

M of coordinates (u,v) and the perpendicular distance p from the point M' of coordinates $(u + du,\ v + dv)$ to this tangent plane. The principal part of $2p$ is then

$$D\ du^2 + 2\ D'\ du\ dv + D''\ dv^2 \qquad (5)$$

where D, D', D'' are functions of the first and second partial derivatives of (x,y,z) and are therefore expressible as functions of u and v. Equation (5) is called the second fundamental form of the surface. E, F, G and D, D', D'' are not independent but must satisfy the so-called Gauss-Codazzi equations. If both forms are given the sur-

desics tangent to the curve at the points s and $s + \Delta s$. This curvature κ_g is measured in the surface and is in general different from the curvature of the curve as measured in space. κ_g is obviously zero for any geodesic so that geodesics play the role of straight lines in the plane.

n-Dimensional applications. The idea of obtaining all intrinsic metric properties of a surface from the first fundamental form was extended by G. F. B. Riemann to a space of n dimensions. In this type of space a point has the coordinates $(u_1, u_2, \ldots, u_n)$ and the arc length is given by

$$ds^2 = \sum_{i,j=1}^{n} g_{ij}(u)\ du_i\ du_j \qquad (6)$$

one tangent plane and one surface normal. A plane in a given direction containing this normal cuts out a normal section of the surface. The curvature of this curve at P is the normal curvature of the surface in the direction of the plane. As the normal plane rotates about the surface normal as an axis, the normal curvature usually changes; it takes on its maximum and its minimum values for two directions and these are at right angles to each other. These are the principal directions at P. If all normal curvatures at P have the same value, P is called an umbilical point. Thus every point on a sphere is umbilical, and a general ellipsoid has four umbilical points. The curves on the surface whose tangents have a principal direction at every point are the lines of curvature of the surface. They are often chosen as the coordinate lines, for in that case $F = 0$, $D' = 0$. Denoting by ρ_1 and ρ_2 the radii of curvature of the principal normal sections at P,

$$K_m = \frac{1}{\rho_1} + \frac{1}{\rho_2}$$

is called the mean curvature of the surface at P and $K = 1/\rho_1\rho_2$ is the total or Gaussian curvature of the surface at P. The value of K is given by

$$K = \frac{DD'' - D'^2}{EG - F^2}$$

As thus defined, K_m and K depend on the enveloping space. However K. F. Gauss showed that K is expressible in terms of E, F, and G and their first and second partial derivatives. Thus K is an intrinsic invariant of the surface; it has the same value for all applicable surfaces. For a plane $K = 0$, for a sphere $K = 1/a^2$, and for a pseudosphere $K = -(1/a^2)$.

Through each point of a surface in a given direction there is one curve whose osculating plane coincides with the normal plane in that direction. These curves are the geodesics of the surface. In a properly restricted neighborhood of a point there is just one geodesic through two points in that neighborhood; it is the shortest distance between the two points. From this it follows that the geodesics of a surface depend only on E, F, and G. The geodesic curvature of a curve on a surface is $d\theta/ds$ where $\Delta\theta$ is the angle between the geo-

Such a space is called Riemannian and the functions $g_{ij}(u)$ are the fundamental quantities for the space. All the intrinsic properties of a surface have been generalized to a Riemannian space though not in a trivial or obvious manner. It was the knowledge of this geometry that enabled Albert Einstein to work out the general theory of relativity, for according to that theory, physical space is four dimensional, the $g_{ij}(u)$ being determined by the matter and energy in the space. See Calculus of tensors; Conformal mapping; Coordinate systems, graphical; Geometry, Riemannian; Projective geometry; Relativity.
[M.S.K.]

Bibliography: L. P. Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces*, 1909; G. Scheffers, *Anwendung der Differential- und Integral-Rechnung auf Geometrie*, 2 vols., 1910–1913; J. L. Synge and A. Schild, *Tensor Calculus*, 1949.

Geometry, euclidean

The word geometry is derived from two Greek words meaning "earth measurement." It seems probable that many of the early discoveries in geometry were motivated by the need to make measurements of distances and areas on the earth. However, euclidean geometry has a broader meaning. It is the chief subject matter of the monumental 13-volume work called *The Elements*, written about 300 B.C. by the Greek mathematician Euclid, who taught and founded a school of geometry at Alexandria. One of the most important milestones in the history of scientific thought, these books of Euclid still occupy an important position in mathematical instruction today.

Geometry, as developed in Euclid, was a systematic body of mathematical knowledge, built by deductive reasoning upon a foundation of three main pillars: (1) definitions of such things as points, lines, planes, angles, circles, and triangles; (2) the assumption of certain geometrical postulates regarded as true but perhaps not self-evident; and (3) the assumption of certain axioms or common notions which were taken to be self-evident truths. The body of Euclid's great work consists of a set of propositions or theorems, each derived

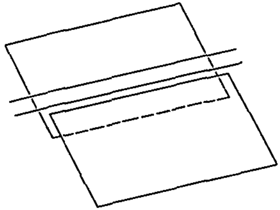


Fig. 1. Parallel lines and planes.

systematically and logically from the definitions, axioms, and postulates of his foundation, and from theorems already proved.

Definitions. Several of the more important definitions are given in this article. Other definitions and explanations are given in separate articles. For example, *see* CIRCLE; POLYGON; TRIANGLE.

Point. A point, in the euclidean definition, is that which has no part. A point, according to most modern authors, is an undefined element in a geometry, such that to each pair of points corresponds a unique set of points called a line In post-euclidean metric geometry, each pair of points A and B determines a (real) number called the distance $\overline{AB}$, which is positive or zero according as A and B are distinct points or the same point.

Line. A line, according to the euclidean definition, is breadthless length, the extremities of a line are points, and a straight line is a line that lies evenly with the points of itself. Three points A, B, and C are said to be collinear if and only if they lie in the same straight line. In modern usage, the finite straight line of Euclid is called a line segment, and the single word line in geometry commonly refers to a euclidean straight line produced indefinitely in both directions. In metric geometry, the point P is said to lie between A and B if $\overline{AP} + \overline{PB} = \overline{AB}$. Three points are collinear if one lies between the other two. A line segment $[AB]$ consists of two distinct points A and B and all the points between them. A line AB consists of two distinct points A and B and all the points collinear with them. A half line, or ray, $A(B)$ consists of the segment $[AB]$ and all points P such that B lies between A and P.

Any two lines (indefinitely extended) are congruent. Two lines having two distinct points in common are coincident; two lines having just one point in common are called intersecting lines. Any two rays (or half lines) are congruent. They are coincident if they have in common the vertex and one other point. Two line segments are congruent if and only if they have the same length.

Skew lines. Two lines that are not coplanar are called skew. Given any two skew lines in space, and a point P not on either, there is a unique line through P that intersects both skew lines. There is a unique plane that contains the first of two given skew lines and that is parallel to the second.

Parallels. Two lines (each extended indefinitely in both its directions) are called parallel if and only if they lie in the same plane but do not intersect. Two planes, or a line and a plane, are called parallel if they do not intersect. There is one and only one line parallel to a given line through a given point not on the line (Euclid's parallel postulate). Two distinct lines parallel to the same line (in space or in a plane) are parallel to each other. Two distinct planes parallel to the same plane are parallel to each other. If each of two parallel planes is intersected by a third plane, the lines of intersection are parallel.

Perpendicular. When two straight lines intersect so that adjacent angles at their point of intersection are equal, these angles are called right angles, and the lines are said to be perpendicular, or normal, or orthogonal to each other. Two skew lines are called perpendicular if one of them intersects at right angles a line parallel to the other. In general a line perpendicular to one of two parallel lines is perpendicular to the other also. In a plane, but not in general in space, two lines each perpendicular to the same line are parallel. One and only one perpendicular can be drawn to a given line through a given point not on the line. The same is true in the plane, but not in space, if the given point is on the line. A line that intersects a plane in a point P is called perpendicular or normal to the plane if it is perpendicular to every line through P that lies in the plane.

The plane is then called perpendicular to the line. A line perpendicular to each of two intersecting lines is perpendicular to their plane and to every line in the plane. The distance from a point to a line or from a point to a plane is measured along the unique perpendicular from the point to the line or plane A line perpendicular to a segment at its midpoint is called a perpendicular bi-

[illegible]

The three perpendicular bisectors of the sides of a triangle meet in a point P called its circumcenter, which is equally distant from all three vertices.

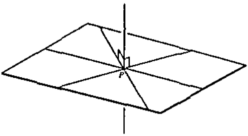


Fig. 2. Normal to a plane.

Congruence. Two geometric figures are said to be congruent to each other if the points in the one figure can be made to correspond in a one-to-one manner with the points of the other figure so that the distance between any two points in the one figure is equal to the distance between corresponding points in the other figure. (This is a post-euclidean version of the definition.)

Postulates. Five postulates, or assumptions, followed Euclid's list of definitions, and may be paraphrased as follows:

P1. A unique straight line can be drawn from any point to any other point.

P2. A finite straight line can be extended continuously in either direction in a straight line.

P3. A circle can be described with any given center and radius.

P4. All right angles are equal.

P5 If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines if produced indefinitely, meet on that side on which the angles are less than two right angles.

This fifth postulate is known as the parallel postulate, and is essentially equivalent to the statement that "a unique line parallel to a given line can be constructed through any point not on the line." For 2000 years many unsuccessful attempts were made to prove that this postulate was a consequence of Euclid's other postulates, definitions, and axioms. Only as recently as the nineteenth century did N. I. Lobachevski (1793–1856), J. Bolyai (1802–1860), and G F. B. Riemann (1826–1866) show that the parallel postulate was independent of the others, by constructing so-called noneuclidean geometries in which the fifth postulate is not valid. *See* GEOMETRY, NONEUCLIDEAN.

Axioms. Certain common notions or axioms were stated by Euclid, and treated as self-evident truths.

A1. Things which are equal to the same thing are equal to each other.

A2. If equals be added to equals, the wholes are equal.

A3. If equals be subtracted from equals, the remainders are equal.

A4. Things which coincide with one another are equal to one another.

A fifth axiom, ascribed by Proclus to Euclid, is believed by Paul Tannery and T. L. Heath to have been added by later writers:

A5. The whole is greater than the part.

This has also been replaced by the statement: The [illegible]

[illegible]

of [illegible]

T[illegible]

are now believed to have been written by other authors. The first six books of Euclid include most of the subject [illegible] geometr[illegible] number [illegible] mensura[illegible] Books 11 and 12 are concerned with solid geometry and mensuration; and

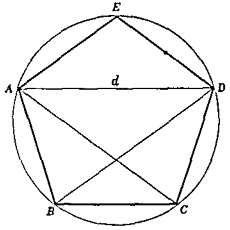


Fig 3. Regular pentagon $ABCDE$ showing diagonal d.

Book 13 with extreme and mean ratio and the five regular solids.

Book 1, after listing a large number of definitions and the postulates and axioms described above, contains 48 propositions including many properties relating to perpendicular and parallel lines, angles, and congruent triangles, and it culminates with the important pythagorean theorem.

Book 2 contains 14 propositions on geometrical algebra, establishing relationships between the areas of certain related squares and rectangles. For example, Proposition 7 is equivalent to the identity $(a+b)^2 = a^2 + b^2 + 2ab$.

Book 3 deals with circles, chords, inscribed angles, and other figures related to the circle; Book 4 with inscribed and circumscribed circles and polygons; Book 5 with ratios; Book 6 with areas and similar triangles.

Book 7 is number theoretical, rather than geometrical, and deals with units, primes, and squares, and with the euclidean algorithm for least common divisor. Book 8 includes continued proportions and cube roots. Book 9 includes the factoring of numbers in continued proportion, a theorem on the infinitude of prime numbers, and a formula for even perfect numbers.

Book 10 includes a thorough treatment of quadratic irrationalities and related incommensurable magnitudes.

Book 11 introduces the fundamentals of three-dimensional geometry, including theorems on planes and lines, perpendiculars and parallels, solid angles, and the volumes of parallelepipedal solids. Book 12 deals with areas and volumes of pyramids and cones and spheres, deriving formulas for the latter by the method of exhaustion due to Eudoxus. Book 13 starts with the division of a segment in extreme and mean ratio—in such a manner that one part of the segment is to the second part as the second part is to the whole. This ratio is the so-called golden ratio of the sides of a regular pentagon to its diagonal d. It is equal to $(\sqrt{5}-1)/2$, its reciprocal is $(\sqrt{5}+1)/2$, and it plays a fundamental role in the study of two of the

regular solids. The latter half of Book 13 is concerned with a study of relations among the volumes and edges and diameters of the five regular solids, often called platonic solids. See GEOMETRY, DIFFERENTIAL; POLYHEDRON; PROJECTIVE GEOMETRY; SOLID (GEOMETRIC). [J.S.F.]

Bibliography: T. L. Heath (trans.), *Euclides, The Thirteen Books of Euclid's Elements*, 1908; S. Maclane, Metric postulates for plane geometry, *Am. Math. Monthly*, 6(7):543–555, 1959.

Geometry, noneuclidean

A system of geometry based upon a set of axioms different from those of the euclidean geometry based on the three-dimensional space of experience. Noneuclidean geometries, especially Riemannian geometry, are useful in mathematical physics. This article will describe some of the basic concepts and axioms of noneuclidean geometries, and then draw some comparisons with euclidean concepts. See GEOMETRY, DIFFERENTIAL; GEOMETRY, EUCLIDEAN; GEOMETRY, RIEMANNIAN; PROJECTIVE GEOMETRY.

The famous names related to hyperbolic (noneuclidean) geometry are J. Bolyai and N. I. Lobachevski. Also A. Cayley, E. Beltrami, K. F. Gauss, and G. F. B. Riemann have given many outstanding contributions to the subject of noneuclidean geometry.

Projective space. Let K denote a given field of elements, called the ground field K. In actual applications, K will be simply the field of real numbers. Consider the class Σ of ordered sets of $(n+1)$-tuples $x = (x^0, x^1, \ldots, x^n)$, where each x^i is an element of the ground field K. The index i is not an exponent but is merely a superscript. In this article, both superscripts and subscripts will be used. If $x = (x^0, x^1, \ldots, x^n)$ and $y = (y^0, y^1, \ldots, y^n)$ are two elements in the class Σ, then the sum $(x + y)$ is defined to be the element $x + y = (x^0 + y^0, x^1 + y^1, \ldots, x^n + y^n)$ in Σ. Also if $x = (x^0, x^1, \ldots, x^n)$ is an element of Σ and if ρ is an element of the ground field K, then the scalar product ρx is defined to be the element $\rho x = (\rho x^0, \rho x^1, \ldots, \rho x^n)$ in Σ.

If $x_1, x_2, \ldots, x_m$ are m elements in Σ and if $\rho_1, \rho_2, \ldots, \rho_m$ is a set of m elements from the ground field K, then the element $x = \rho_1 x_1 + \rho_2 x_2 + \cdots + \rho_m x_m$ of Σ is called a linear combination of the elements $x_1, x_2, \ldots, x_m$ of Σ.

The class Σ^1 is a proper subset of Σ which is composed of all ordered sets of $(n+1)$-tuples $x = (x^0, x^1, \ldots, x^n)$ of Σ such that at least one x^i of the $(n+1)$ elements $x^0, x^1, \ldots, x^n$ of the ground field K is not zero. Two elements $x = (x^0, x^1, \ldots, x^n)$ and $y = (y^0, y^1, \ldots, y^n)$, of the class Σ^1, are said to be equivalent if and only if there exists an element $\rho \neq 0$, of the ground field K, such that $y = \rho x$. This establishes an equivalence relation among the elements $x = (x^0, x^1, \ldots, x^n)$ of the class Σ^1. The equivalence classes in Σ^1, formed relative to this equivalence relation, are called points P. The collection of all these points P forms a projective space S_n of n dimensions over the ground field K.

A point P in S_n is a class in Σ^1 of all ordered sets of $(n+1)$-tuples equivalent to a given set of $(n+1)$-tuples, namely $x = (x^0, x^1, \ldots, x^n)$. Each of these ordered sets of $(n+1)$-tuples is in the equivalence class called the point P, and is said to be a set of homogeneous point coordinates for the point P. Consequently if $x = (x^0, x^1, \ldots, x^n)$ is a set of homogeneous point coordinates of a point P in S_n, any other set of homogeneous point coordinates for the same point P is $\rho x = (\rho x^0, \rho x^1, \ldots, \rho x^n)$, where $\rho \neq 0$ is an arbitrary element of the ground field K.

A collection of $(p+1)$ points $x_0, x_1, \ldots, x_p$ in S_n is said to be linearly dependent if and only if at least one of them is a linear combination of the remaining ones such that at least one of the scalar coefficients is not zero. Otherwise they are said to be linearly independent. Clearly, if $(p+1)$ points $x_0, x_1, \ldots, x_p$ in S_n are linearly independent, then $0 \leqq p \leqq n$.

Let a collection of $(p+1)$ linearly independent points $x_0, x_1, \ldots, x_p$ in S_n be given. Then $0 \leqq p \leqq n$. The S_p determined by them is composed of all points x in S_n such that

$$x = y^0 x_0 + y^1 x_1 + \cdots + y^p x_p \qquad (1)$$

where at least one of the scalar multiples $y^0, y^1, \ldots, y^p$ is not zero. This S_p is a projective space of p dimensions over the ground field K, and is a projective subspace of S_n. A set of homogeneous point coordinates for an arbitrary point P in this S_p is $(y^0, y^1, \ldots, y^p)$.

Define S_{-1} to be a vacuous collection of points P in S_n. Such an S_{-1} is called a projective subspace of S_n of dimension -1. Evidently every S_0 is a point P and conversely. A line is an S_1, a plane is an S_2, a three-space is an S_3, and a p-space is an S_p. In particular, a hyperplane is an S_{n-1}.

Dual coordinates. An S_p is a locus of points x whose homogeneous point coordinates $x = (x^0, x^1, \ldots, x^n)$ satisfy a system of $(n-p)$ linear homogeneous equations

$$\sum_{j=0}^{n} u_{ij} x^j = 0 \quad (i = 1, 2, \ldots, n-p) \qquad (2)$$

such that the rank of the coefficient matrix (u_{ij}) is $(n-p)$.

In particular, a hyperplane S_{n-1} is a locus of points x whose homogeneous point coordinates $x = (x^0, x^1, \ldots, x^n)$ satisfy a single linear homogeneous equation

$$ux = \sum_{i=0}^{n} u_i x^i = u_0 x^0 + u_1 x^1 + \cdots + u_n x^n = 0 \qquad (3)$$

where at least one of the $(n+1)$ elements $u_0, u_1, \ldots, u_n$, of the ground field K, is not zero.

Because an ordered set of $(n+1)$-tuples $u = (u_0, u_1, \ldots, u_n)$ of this type uniquely defines an S_{n-1}, it is called a set of homogeneous hyperplane coordinates, for the particular S_{n-1}. If u

$(u_0,u_1, \ldots ,u_n)$ is a set of homogeneous hyperplane coordinates of an S_{n-1}, any other set of homogeneous hyperplane coordinates for the same S_{n-1} is $\sigma u = (\sigma u_0,\sigma u_1, \ldots ,\sigma u_n)$, where $\sigma \neq 0$ is an arbitrary element of the ground field K.

In the projective space S_n of n dimensions, a point P and a hyperplane S_{n-1} are termed dual objects. Thus, homogeneous point coordinates $x = (x^0,x^1, \ldots ,x^n)$ and homogeneous hyperplane coordinates $u = (u_0,u_1, \ldots ,u_n)$ are said to be dual systems of coordinates.

A point P with the homogeneous point coordinates $x = (x^0,x^1, \ldots ,x^n)$ is on the hyperplane S_{n-1} with the homogeneous hyperplane coordinates $u = (u_0,u_1, \ldots ,u_n)$ if and only if the condition of Eq. (3) is satisfied.

By mathematical induction, it follows that the dual of a projective subspace S_p of p dimensions is a projective subspace S_{n-p-1} of $(n-p-1)$ dimensions. For example, the dual of S_{-1} is S_n, the dual of an S_0 is an S_{n-1}, and the dual of an S_1 is an S_{n-2}.

The projective subspace S_p is said to be contained in the projective subspace S_q, or the projective subspace S_q contains the projective subspace S_p, if and only if every point of S_p is a point of S_q. In particular, S_{-1} is contained in every S_q, and S_n contains every S_p. This relation is written as $S_p \subset S_q$, or $S_q \supset S_p$.

The dual of the relation $S_p \subset S_q$ is the dual relation $S_{n-p-1} \supset S_{n-q-1}$.

Principle of duality. For every theorem concerning the S_ps and the relations $S_p \subset S_q$, there is obtained a dual theorem wherein each S_p is replaced by the dual object S_{n-p-1}, and each relation $S_p \subset S_q$ is replaced by the dual relation $S_{n-p-1} \supset S_{n-q-1}$.

If S_p and S_q are any two projective subspaces of the projective space S_n, the largest projective subspace contained in both S_p and S_q is denoted by $S_p \cap S_q$, and the smallest projective subspace containing both S_p and S_q is denoted by $S_p \cup S_q$. Clearly, $S_p \cap S_q$ is the intersection or meet, and $S_p \cup S_q$ is the union or join of S_p and S_q. *See* SET THEORY.

The two relations $S_p \cap S_q$ and $S_p \cup S_q$ are dual. Concerning the inclusion relation $S_p \subset S_q$, the projective subspaces of the projective space S_n form a

dimension theorem may be stated in the form

$$\dim (S_p \cap S_q) + \dim (S_p \cup S_q) = \dim (S_p) + \dim (S_q) \tag{4}$$

Projective group. A collineation T of the projective space S_n is a one-to-one correspondence between the points of S_n, of the form

$$\rho \bar{x}^i = \sum_{j=0}^{n} a_j{}^i x^j \quad (i = 0, 1, \ldots, n) \tag{5}$$

where the rank of the coefficient mat
$(n+1)$, and $\rho \neq 0$ is an arbitrary cons
portionality. Of course, all elements ar
original ground field K.

A collineation T is not only a one-to-one correspondence between the points S_0 of S_n, but also a one-to-one correspondence between the S_ps of S_n; that is, under T, any S_p is carried into one and only one $\bar{S}_p$, and conversely.

The set of collineations T of S_n forms the collineation group G of S_n. It is composed of $n(n+2)$ essential parameters.

The fundamental theorem of projective geometry may be stated in the following form: there is a collineation T which carries a given set of $(p+1)$ linearly independent points into a prescribed set of $(p+1)$ linearly independent points. Moreover, if $p = n$, then T is uniquely determined.

A correlation Γ of the projective space S_n is a one-to-one correspondence between the points S_0 and the hyperplanes S_{n-1} of S_n, of the form

$$\sigma \bar{u}_i = \sum_{j=0}^{n} b_{ij} x^j \quad (i = 0, 1, \ldots, n) \tag{6}$$

where the rank of the coefficient matrix (b_{ij}) is $(n+1)$, and $\sigma \neq 0$ is an arbitrary constant of proportionality.

The set C of correlations Γ of S_n is composed of $n(n+2)$ essential parameters. In general, C is not a group.

The total projective group G^* of the projective space S_n is composed of collineations T and correlations Γ. It is a mixed group G^* of $n(n+2)$ essential parameters. Obviously, the collineation group G is a subgroup of the total projective group G^*.

Projective geometry consists of the qualitative and quantitative invariants of the total projective group G^*.

Cross ratio. Let S_{r-1} and S_{r+1} be two fixed projective subspaces of S_n of dimensions $(r-1)$ and $(r+1)$, respectively, such that S_{r-1} is contained in S_{r+1}. Evidently $0 \leqq r \leqq n-1$. A pencil is composed of all the S_rs that contain the given S_{r-1} and that are contained in the given S_{r+1}.

For example, when $r = 0$, one obtains a pencil of points, all of which are in a fixed line S_1. Similarly when $r = n-1$, there is defined a pencil of hyperplanes S_{n-1}, all of which contain a fixed projective subspace S_{n-2} of $(n-2)$ dimensions.

A pencil of elements P is a projective space of one dimension. Therefore, the elements P of a pencil are defined by a system of homogeneous coordinates (ρ,τ), where at least one of the elements ρ and τ of the ground field K is not zero. If it is assumed that τ is always one, then ρ is said to be the nonhomogeneous coordinate of the element P which is not the element at infinity $(1,0)$. The element of infinity $(1,0)$ in this system of coordinates is denoted by the symbol ∞.

If $(\rho_1,\tau$), (ρ_3,τ_3), (ρ_4,τ_4) are the homogeneo es of four distinct elements
P_1, P_2, P_3 ross ratio R is defined to be
the numeri

$$\mathrm{R}(P_1P_2 \quad \frac{(\rho_2\tau_3 - \rho_3\tau_2)}{\tau_1 - \rho_1\tau_4)} \tag{7}$$

This is the fundamental projective invariant. In nonhomogeneous coordinates, this is

$$\mathrm{R}(P_1P_2,P_3P_4) = \frac{(\rho_3 - \rho_1)(\rho_2 - \rho_4)}{(\rho_2 - \rho_3)(\rho_4 - \rho_1)} \tag{8}$$

In particular,

$$\mathrm{R}(\infty P_3,P_1P_2) = \frac{\rho_3 - \rho_2}{\rho_3 - \rho_1} \tag{9}$$

Also

$$\mathrm{R}(\infty 0,P_1P_2) = \frac{\rho_2}{\rho_1} \tag{10}$$

Of importance in projective geometry is the concept of a harmonic set of elements of a pencil provided that the ground field K is not of characteristic two. Four elements P_1, P_2, P_3, P_4, of a pencil, are said to form a harmonic set of elements if and only if $\mathrm{R}(P_1P_2, P_3P_4) = -1$.

Quadrics. Consider a correspondence Γ of S_n in which every point P is converted into a single hyperplane S_{n-1}, or S_n. It is assumed that Γ carries every line S_1 into a single S_{n-2}, or S_{n-1}, or S_n. Finally, it is supposed that if Γ carries a point P into an S_p and if $\bar{P}$ is any point of this S_p, then Γ converts the point $\bar{P}$ into an S_q which passes through the original point P. Such a correspondence Γ is called a polarity Γ.

In homogeneous coordinates, a polarity Γ is given by the equations

$$\sigma \bar{u}_i = \sum_{j=0}^{n} g_{ij}x^j \quad (i = 0, 1, \ldots, n) \tag{11}$$

where the matrix (g_{ij}) is symmetric, that is, $g_{ij} = g_{ji}$, and where its rank is $(r + 1)$ for which $0 \leqq r \leqq n$. If $0 \leqq r < n$, the polarity Γ is said to be singular. Otherwise Γ is said to be nonsingular.

In the homogeneous point coordinates, the polarity Γ is given by the single equation

$$\sum_{i,j=0}^{n} g_{ij}x^i\bar{x}^j = 0 \tag{12}$$

The dual of a polarity Γ is also a polarity Γ^*. Such a polarity Γ^* is given in homogeneous coordinates by the equations

$$\rho \bar{x}^i = \sum_{j=0}^{n} g^{ij}u_j \quad (i = 0, 1, \ldots, n) \tag{13}$$

where the matrix (g^{ij}) is symmetric, that is $g^{ij} = g^{ji}$, and where its rank is $(r + 1)$ for which $0 \leqq r \leqq n$.

In homogeneous hyperplane coordinates, the polarity Γ^* is given by the single equation

$$\sum_{i,j=0}^{n} g^{ij}\bar{u}_i\bar{u}_j = 0 \tag{14}$$

If the polarity Γ is nonsingular, then the dual Γ^* is the original polarity Γ. In that event, the two matrices (g_{ij}) and (g^{ij}) can be considered to be inverse matrices. Thus a nondegenerate polarity can be given by Eqs. (11), (12), (13), or (14).

A hyperplane element (P,π) is composed of a point P and a hyperplane π of dimension $(n - 1)$ which passes through the point P. A quadric Q is a locus of hyperplane elements (P,π) such that under a given polarity Γ or Γ^*, the point P is transformed into the hyperplane π, or the hyperplane π is carried into the point P.

The polarity Γ or Γ^* is said to be a polarity relative to the corresponding quadric Q.

In homogeneous point coordinates, the equation of a quadric Q is

$$\sum_{i,j=0}^{n} g_{ij}x^ix^j = 0 \tag{15}$$

In homogeneous hyperplane coordinates, the equation of a quadric Q is

$$\sum_{i,j=0}^{n} g^{ij}u_iu_j = 0 \tag{16}$$

Euclidean and noneuclidean geometries. Consider a real projective space S_n. That is, S_n is defined over the real number system K. Sometimes it is convenient to consider that S_n is immersed in a complex projective space S_n^*, which is defined over the complex number system K^*.

Let $k \neq 0$ be either a positive real number or infinite (that is, $1/k = 0$), or else a pure imaginary number of the form $k = il$, where $i^2 = -1$ and l is a positive real number. Consider the fundamental quadric Σ whose homogeneous point equation is

$$f(x,x) = (kx^0)^2 + (x^1)^2 + (x^2)^2 + \cdots + (x^n)^2 = 0 \tag{17}$$

where the superscripts exterior to the parentheses denote exponents. The homogeneous hyperplane equation of this fundamental quadratic Σ is

$$F(u,u) = \left(\frac{u_0}{k}\right)^2 + u_1^2 + u_2^2 + \cdots + u_n^2 = 0 \tag{18}$$

The set of all collineations T of S_n which carry this fundamental quadric Σ into itself is a group $G(k)$ of $n(n + 1)/2$ essential parameters. When k is a positive real number, this is the elliptic group G_E of elliptic geometry. When $1/k = 0$, this is the euclidean group G_P of euclidean geometry. Finally when $k = il$ where $i^2 = -1$ and l is a positive real number, this is the hyperbolic group G_H of hyperbolic geometry.

The study of the qualitative and quantitative invariants of these three groups G_E, G_P, and G_H constitutes the three subjects of elliptic, euclidean, and hyperbolic geometries. The elliptic and hyperbolic geometries are usually referred to as the noneuclidean geometries.

In elliptic geometry, two distinct lines contained in a single plane always meet in a single point. Therefore, if L is a fixed line and if P is a

point not in L, there cannot be a line M passing through this point P which is parallel to the given line L.

In euclidean geometry, the ideal hyperplane π_∞ or the hyperplane π_∞ at infinity is the one whose point equation is $x^0 = 0$, or whose hyperplane equations are $u_1 = 0,\ u_2 = 0, \ldots, u_n = 0$. The proper S_ps for $p = -1, 0, 1, \ldots, n$ are those which are not contained in the ideal hyperplane π_∞. The improper S_ps for $p = 0, 1, \ldots, n-1$ are those which are contained in the ideal hyperplane π_∞. In euclidean geometry, only the proper S_ps are studied.

If a proper S_p and a distinct proper S_q intersect in an improper S_r, then S_p and S_q are said to be parallel. Thus, two distinct lines in euclidean space are said to be parallel if and only if they intersect in an ideal point.

From the preceding discussion, Euclid's fifth parallel postulate is an easy consequence. That is, if in euclidean space, L is a fixed line and if P is a fixed point not on L, there is one and only one line M parallel to L and passing through P.

In hyperbolic geometry, the points and the tangent S_ps for $p = 0, 1, 2, \ldots, n-1$ of the fundamental quadric Σ given by Eqs. (17) or (18) are said to be ordinary improper or ordinary infinite. The S_ps for $p = 0, 1, 2, \ldots, n-1$, which are in the exterior of this quadric Σ, are said to be ultraimproper or ultrainfinite. The proper points in hyperbolic geometry are those which are in the interior of this quadric Σ. The proper S_ps for $p = 1, 2, \ldots, n-1, n$ are those which contain proper points and are considered to be sets of these proper points.

If a proper S and a distinct . . .

or ultraparallel.

In hyperbolic geometry, if L is a fixed proper line and if P is a given proper point not on this line L, then there are two distinct proper lines M_1 and M_2 passing through P which are ordinary lines parallel to L. Also passing through P, there is an infinite number of proper lines M which are ultraparallel to L. These lines M belong to the flat pencil with vertex at P and determined by the lines M_1 and M_2.

Distance. Let P and Q be two distinct points given by the homogeneous point coordinates $x = (x^0, x^1, \ldots, x^n)$ and $y = (y^0, y^1, \ldots, y^n)$. In euclidean and hyperbolic geometries, it is understood that P and Q are proper points. The line L determined by the two points P and Q intersects the fundamental quadric Σ in two distinct points P_∞ and Q_∞. The distance $s = s(P,Q)$ between these two points P and Q is defined by the formula

$$s = s(P,Q) = \frac{k}{2i} \log \mathrm{R}(PQ,P_\infty Q_\infty) \qquad (19)$$

It is understood that $s = s(P,Q)$ is a real nonnegative number. Also it is assumed in elliptic geometry that $0 \leq s/k \leq \pi$.

Let $f(x,y)$ denote the expression

$$f(x,y) = k^2x^0y^0 + x^1y^1 + x^2y^2 + \cdots + x^ny^n \qquad (20)$$

The two points P and Q are polar reciprocal or orthogonal relative to the fundamental quadric Σ if and only if $f(x,y) = 0$. The point P is on Σ if and only if $f(x,x) = 0$.

A point R whose homogeneous point coordinates are $z = (z^0, z^1, \ldots, z^n)$ is on the line L determined by the two points P and Q if and only if a number ρ exists such that

$$z^0 = x^0 + \rho y^0, \quad z^1 = x^1 + \rho y^1, \quad \ldots, \quad z^n = x^n + \rho y^n \qquad (21)$$

This point R is on the fundamental quadric Σ whose point equation is Eq. (17) if and only if ρ satisfies the quadratic equation

$$f(x,x) + 2\rho f(x,y) + \rho^2 f(y,y) = 0 \qquad (22)$$

Because the two points P and Q are distinct, this will have two distinct roots ρ_1 and ρ_2. The two points P_∞ and Q_∞ are on the line L whose homogeneous point equations are Eq. (21), corresponding to the two distinct roots ρ_1 and ρ_2.

In elliptic and euclidean geometries, these two points P_∞ and Q_∞ are conjugate-imaginary. In hyperbolic geometry, they are real.

From Eqs. (21) and (22), it is seen that

$$\mathrm{R}(PQ,P_\infty Q_\infty) = \frac{\rho_1}{\rho_2} = \frac{-f(x,y) - \sqrt{f^2(x,y) - f(x,x)f(y,y)}}{-f(x,y) + \sqrt{f^2(x,y) - f(x,x)f(y,y)}} \qquad (23)$$

Then

$$\mathrm{R}(PQ,P_\infty Q_\infty) = \frac{[f(x,y) + \sqrt{f^2(x,y) - f(x,x)f(y,y)}]^2}{f(x,x)f(y,y)} \qquad (24)$$

Consequently the distance $s(P,Q)$ is given by the formula

$$s(P,Q) = \frac{k}{i} \log \frac{f(x,y) + \sqrt{f^2(x,y) - f(x,x)f(y,y)}}{\sqrt{f(x,x)}\,\sqrt{f(y,y)}} \qquad (25)$$

Set $s = s(P,Q)$. Then s is given by the equations

$$\cos\frac{s}{k} = \frac{f(x,y)}{\sqrt{f(x,x)}\,\sqrt{f(y,y)}} \qquad \sin\frac{s}{k} = \frac{\sqrt{f^2(x,y) - f(x,x)f(y,y)}}{i\sqrt{f(x,x)}\,\sqrt{f(y,y)}} \qquad (26)$$

Of course, from Eq. (19), this distance $s = s(P,Q)$ is invariant under each of the groups G_E, G_P, G_H of elliptic, euclidean, and hyperbolic geometries.

By Eq. (20), the Eqs. (26) may be written in the forms shown in Eqs. (27), in which $\epsilon = +1$ or $\epsilon = -1$, according to whether the geometry is elliptic (including the euclidean case) or hyperbolic.

In hyperbolic geometry, $k = il$ where $i^2 = -1$ and l is a positive real number. Equations (27) can be written in the forms shown in Eqs. (28).

$$\cos\frac{s}{k} = \frac{\epsilon(k^2x^0y^0 + x^1y^1 + x^2y^2 + \cdots + x^ny^n)}{\sqrt{[(kx^0)^2 + (x^1)^2 + \cdots + (x^n)^2][(ky^0)^2 + (y^1)^2 + \cdots + (y^n)^2]}}$$

$$\sin\frac{s}{k} = \sqrt{\frac{\epsilon\left[k^2\sum_{i=1}^{n}\begin{vmatrix} x^0 & x^i \\ y^0 & y^i\end{vmatrix}^2 + \frac{1}{2}\sum_{i,j=1}^{n}\begin{vmatrix} x^i & x^j \\ y^i & y^j\end{vmatrix}^2\right]}{[(kx^0)^2 + (x^1)^2 + \cdots + (x^n)^2][(ky^0)^2 + (y^1)^2 + \cdots + (y^n)^2]}} \quad (27)$$

$$\cosh\frac{s}{l} = \frac{l^2x^0y^0 - x^1y^1 - x^2y^2 - \cdots - x^ny^n}{\sqrt{[(lx^0)^2 - (x^1)^2 - \cdots - (x^n)^2][(ly^0)^2 - (y^1)^2 - \cdots - (y^n)^2]}}$$

$$\sinh\frac{s}{l} = \sqrt{\frac{l^2\sum_{i=1}^{n}\begin{vmatrix} x^0 & x^i \\ y^0 & y^i\end{vmatrix}^2 - \frac{1}{2}\sum_{i,j=1}^{n}\begin{vmatrix} x^i & x^j \\ y^i & y^j\end{vmatrix}^2}{[(lx^0)^2 - (x^1)^2 - \cdots - (x^n)^2][(ly^0)^2 - (y^1)^2 - \cdots - (y^n)^2]}} \quad (28)$$

Return to the general case of Eqs (27). Let $x = (x^0, x^1, \ldots, x^n)$ denote affine coordinates in a euclidean space of $(n+1)$ dimensions. Define the special quadric Σ^* by the equation

$$f(x,x) = (kx^0)^2 + (x^1)^2 + \cdots + (x^n)^2 = k^2 \quad (29)$$

This is a central quadric Σ^*. Each of the two noneuclidean geometries can be visualized as the geometry on this quadric Σ^* in which diametrically opposite points are identified. In particular, for elliptic geometry, Σ^* can be considered a sphere.

On this quadric Σ^*, Eqs. (27) can be written in the forms

$$\cos\frac{s}{k} = x^0y^0 + \frac{1}{k^2}(x^1y^1 + x^2y^2 + \cdots + x^ny^n)$$

$$k\sin\frac{s}{k} = \sqrt{\sum_{i=1}^{n}\begin{vmatrix} x^0 & x^i \\ y^0 & y^i\end{vmatrix}^2 + \frac{1}{2k^2}\sum_{i,j=1}^{n}\begin{vmatrix} x^i & x^j \\ y^i & y^j\end{vmatrix}^2} \quad (30)$$

In the case of elliptic geometry, the quadric Σ^* can be considered to be a sphere. Here k is a positive real number and the distance s, such that $0 \leqq s/k \leqq \pi$, is given by the preceding equations.

Euclidean geometry can be considered to be the limiting case of either elliptic or hyperbolic geometry as k becomes infinite. In this case, one can always regard x^0 as unity. Then from Eq. (30), the distance formula $s = s(P,Q)$ for euclidean geometry is

$$s = s(P,Q) = \sqrt{(x^1 - y^1)^2 + (x^2 - y^2)^2 + \cdots + (x^n - y^n)^2} \quad (31)$$

In this case, coordinates $x = (x^1, x^2, \ldots, x^n)$ of a point P are said to be rectangular or cartesian.

The final case is that of hyperbolic geometry. Here $k = il$ where $i^2 = -1$ and l is a positive real number. From Eq. (30), the distance formula $s = s(P,Q)$ is given by the equations

$$\cosh\frac{s}{l} = x^0y^0 - \frac{1}{l^2}(x^1y^1 + x^2y^2 + \cdots + x^ny^n)$$

$$l\sinh\frac{s}{l} = \sqrt{\sum_{i=1}^{n}\begin{vmatrix} x^0 & x^i \\ y^0 & y^i\end{vmatrix}^2 - \frac{1}{2l^2}\sum_{i,j=1}^{n}\begin{vmatrix} x^i & x^j \\ y^i & y^j\end{vmatrix}^2} \quad (32)$$

In each of the three geometries, it is assumed that the relationship $s = s(P,Q)$ is zero if and only if the two points P and Q are identical.

Each of the three geometries is an abstract metric space. That is,

(i) $s(P,Q) \geqq 0$, and $s(P,Q) = 0$ if and only if $P = Q$
(ii) $s(P,Q) = s(Q,P)$
(iii) $s(P,Q) + s(Q,R) \geqq s(P,R)$

The condition (i) is that of positive definiteness The condition (ii) is that of symmetry. Condition (iii) is the well-known triangular inequality.

Angle. By dualizing the concept of distance $s = s(P,Q)$, the concept of angle $\theta = \theta(\pi,\sigma)$ between two hyperplanes π and σ is obtained. Let π and σ be two distinct hyperplanes which are given by the homogeneous hyperplane coordinates $u = (u_0, u_1, \ldots, u_n)$ and $v = (v_0, v_1, \ldots, v_n)$. In euclidean and hyperbolic geometries, it is understood that π and σ are proper hyperplanes which determine a pencil λ. In this pencil λ there are two distinct hyperplanes π_∞ and σ_∞ which are tangent to the fundamental quadric Σ. The angle $\theta = \theta(\pi,\sigma)$, where $0 \leqq \theta \leqq \pi$, between these two hyperplanes π and σ is defined by the formula

$$\theta = \theta(\pi,\sigma) = \frac{l}{2i}\log R(\pi\sigma,\pi_\infty\sigma_\infty) \quad (33)$$

This angle $\theta = \theta(\pi,\sigma)$ is given by Eqs. (26), in which $\epsilon = +1$ or $\epsilon = -1$ according to whether the geometry is elliptic (including the euclidean case) or hyperbolic.

It is evident that when $\epsilon = +1$ and k becomes infinite, the preceding formulas become those of the angle θ between the two hyperplanes [illegible] in a euclidean space of n dimensions.

The two hyperplanes π and σ are [illegible] orthogonal if and only if [illegible] and σ_2 be two distinct hyperplanes which [illegible] in a line L. Then L is said to be orthogonal [illegible] hyperplane π if σ_1 and σ_2 are each orthogonal [illegible] If L and M are two distinct lines [illegible] through a point P and if M is contained [illegible] plane π orthogonal to L, then L and [illegible] orthogonal.

$$\cos\theta = \frac{\epsilon\left(\dfrac{u_0v_0}{k^2} + u_1v_1 + u_2v_2 + \cdots + u_nv_n\right)}{\sqrt{\left(\dfrac{u_0^2}{k^2} + u_1^2 + \cdots + u_n^2\right)\left(\dfrac{v_0^2}{k^2} + v_1^2 + \cdots + v_n^2\right)}}$$

$$\sin\theta = \sqrt{\frac{\epsilon\left[\dfrac{1}{k^2}\sum_{i=1}^{n}\begin{vmatrix} u_0 & u_i \\ v_0 & v_i \end{vmatrix}^2 + \dfrac{1}{2}\sum_{i,j=1}^{n}\begin{vmatrix} u_i & u_j \\ v_i & v_j \end{vmatrix}^2\right]}{\left(\dfrac{u_0^2}{k^2} + u_1^2 + \cdots + u_n^2\right)\left(\dfrac{v_0^2}{k^2} + v_1^2 + \cdots + v_n^2\right)}} \tag{34}$$

If π is a fixed hyperplane and if P is a point not in π, then there is one and only one line L passing through the point P orthogonal to the hyperplane π.

Differential of arc length. If P and Q are two nearby points on the quadric Σ^* defined by the two sets of coordinates $(\bar{x}^0,\bar{x}^1, \ldots ,\bar{x}^n)$ and $(\bar{x}^0 + d\bar{x}^0,\ \bar{x}^1 + d\bar{x}^1, \ldots ,\ \bar{x}^n + d\bar{x}^n)$, then by Eq. (30), the square of the differential ds of arc length s between the points P and Q is

$$ds^2 = \sum_{i=1}^{n} (\bar{x}^0\, d\bar{x}^i - \bar{x}^i\, d\bar{x}_0)^2 + \frac{1}{2k^2}\sum_{i,j=1}^{n} (\bar{x}^i\, d\bar{x}^j - \bar{x}^j\, d\bar{x}^i)^2 \tag{35}$$

For the point P on this quadric Σ^*, introduce a new set of coordinates $(x^1,x^2, \ldots ,x^n)$ defined by

$$x^i = \frac{2\bar{x}^i}{1 + \bar{x}^0} \quad (i = 1, 2, \ldots , n) \tag{36}$$

where it is understood that $\bar{x}^0 \neq -1$.

In this new set of coordinates, the quadric Σ^* is given by the equation

$$(x^1)^2 + (x^2)^2 + \cdots + (x^n)^2 + 4k^2 = \frac{8k^2}{1 + \bar{x}^0} \tag{37}$$

In the euclidean and noneuclidean geometries, the square of the differential ds of arc length s is given by the equation

$$ds^2 = \frac{(dx^1)^2 + (dx^2)^2 + \cdots + (dx^n)^2}{\left[1 + \dfrac{(x^1)^2 + (x^2)^2 + \cdots + (x^n)^2}{4k^2}\right]^2} \tag{38}$$

As k approaches infinity, the differential $d\sigma$ of arc length σ of euclidean geometry is obtained. If ds represents the differential of arc length s in elliptic or hyperbolic geometry, then

$$ds = \rho\, d\sigma \tag{39}$$

where the scale ρ is

$$\rho = \frac{1}{1 + \dfrac{(x^1)^2 + (x^2)^2 + \cdots + (x^n)^2}{4k^2}} \tag{40}$$

Thus, *each of the noneuclidean geometries may be visualized as a conformal image of euclidean geometry.*

Each of these three geometries is a special case of Riemannian geometry. Let the $g_{ij}(x)$ be $n(n + 1)/2$ real functions which are continuous and possess continuous partial derivatives of as high an order as is necessary in a certain region of real n-dimensional space for which the coordinates of a point P are $x = (x^1,x^2, \ldots ,x^n)$. The g_{ij}s are symmetric; that is, $g_{ij} = g_{ji}$, for all i, $j = 1, 2, \ldots , n$. The quadratic form

$$\sum_{i,j=1}^{n} g_{ij}\lambda^i\lambda^j$$

is assumed to be positive definite, that is

$$\sum_{i,j=1}^{n} g_{ij}\lambda^i\lambda^j$$

is nonnegative; it is zero if and only if $\lambda^i = 0$, for all $i = 1, 2, \ldots , n$. Then the Riemannian space V_n is one for which the square of the differential ds of arc length s between two nearby points P and Q is

$$ds^2 = \sum_{i,j=1}^{n} g_{ij}\, dx^i\, dx^j \tag{41}$$

For this Riemannian space V_n, the Christoffel symbols of the first kind are

$$\Gamma_{ij,k} = \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\right) \tag{42}$$

for $i, j, k = 1, 2, \ldots , n$. The Christoffel symbols of the second kind are

$$\Gamma_{jk}{}^i = \sum_{l=1}^{n} g^{il}\Gamma_{jk;l} \tag{43}$$

for $i, j, k = 1, 2, \ldots , n$. Of course, (g^{ij}) is the inverse matrix of (g_{ij}). The Christoffel symbols of the second kind are also called the affine connections of V_n.

The vector geodesic curvature κ^i of a curve C in V_n is

$$\kappa^i = \frac{d^2x^i}{ds^2} + \sum_{j,k=1}^{n} \Gamma_{jk}{}^i\frac{dx^j}{ds}\frac{dx^k}{ds} \tag{44}$$

A curve C of V_n is said to be a geodesic if and only if $\kappa^i = 0$ for $i = 1, 2, \ldots , n$ at every point P of C.

Each of the three spaces already discussed is a Riemannian space. The geodesics of any one such space are the lines of the space.

A Riemannian space of constant curvature *is applicable to* (that is, can be mapped isometrically into) elliptic, euclidean, or hyperbolic space. In

this case, this constant curvature is equal to the Gaussian curvature G. From Eqs. (35) or (38), this constant Gaussian curvature G is

$$G = 1/k^2 \tag{45}$$

For elliptic space, G is of positive constant curvature. For euclidean space, G is identically zero. Finally for hyperbolic space, G is of negative constant curvature.

In each of these three geometries, consider a geodesic triangle. This is formed by three points P, Q, R, not all on one geodesic, and the three geodesics passing through every two of the three points P, Q, R. If A, B, C are the angles of this geodesic triangle, and if T denotes its area, then

$$A + B + C - \pi = T/k^2 \tag{46}$$

Thus according to whether the geometry is elliptic, euclidean, or hyperbolic, the sum of the angles of a geodesic triangle is greater than π, equal to π, or less than π.

Elliptic geometry of two dimensions may be represented upon a sphere in euclidean space of three dimensions. On the other hand, hyperbolic geometry of two dimensions can be depicted on a pseudosphere in euclidean space of three dimensions. The pseudosphere is obtained by revolving the tractrix about its asymptote. [J.D.CL.]

Bibliography: J. L. Coolidge, *The Elements of Non-Euclidean Geometry*, 1909; J. L. Coolidge, *A History of Geometrical Methods*, 1940; L. P. Eisenhart, *Non-Riemannian Geometry*, 1927; L. P. Eisenhart, *Riemannian Geometry*, 1949; E. Kasner and J. Newman, *Mathematics and the Imagination*, 1940; O. Veblen and J. W. Young, *Projective Geometry*, vols. 1–2, 1910; C. E. Weatherburn, *Differential Geometry of Three Dimensions*, vols. 1–2, 1930

Geometry, Riemannian

The geometry of an N-dimensional space which bears a coordinate system (x) and has associated with it and the given coordinate system a set of N^2 functions $g_{ab}(x)$ which are involved in the determination of certain fundamental geometric magnitudes, including lengths of arcs in the space, lengths of vectors, angles between vectors, the measures of parts of the space (areas, volumes, and hypervolumes), and curvatures. The arc length s, for example, of a parameterized arc $x^a = x^a(t)$ of class C' is given by the formula

$$\Delta s = s(t_2) - s(t_1) = \int_{t_1}^{t_2} \sqrt{g_{ab}x'^a x'^b}\, dt$$

Here, the prime (′) indicates differentiation with respect to t so that $x'^a = dx^a/dt$ and Einstein's summation convention is in force. The quantities g_{ab} are, as indicated, functions of the N coordinate variables $x^1, x^2, \ldots, x^N$ denoted briefly by x. The general subject may be classified into subspecies by means of the value of N and the conditions imposed on the fundamental quantities g_{ab} either directly or through derived quantities.

Riemannian geometry was initiated in 1854 by Bernhard Riemann (1826–1866).

The importance of this subject is due to several facts. First, it includes both euclidean geometry and the geometry of surfaces in euclidean space and, in addition, provides their most natural extension, a generalization of great richness and scope. Second, it provides a geometrical realization or application for certain of the major abstractions of tensor analysis and together with that discipline, with which it merges, forms the pattern and much of the motivation for general relativity. Third, most of the work on the incorporation of electricity and magnetism into the mold of general relativity (unified field theory) almost necessarily involves a Riemannian geometry of one type or another (nonsymmetric g_{ab}, symmetric g_{ab} with $N = 5$, g_{abs} in projective coordinates). Fourth, Riemannian geometry provides the steppingstone to a variety of generalized differential geometries.

Finally, in the absence of contrary evidence, the success of the general theory of relativity suggests that actual physical space may correspond more closely to a noneuclidean Riemannian geometry than to the euclidean variety.

Riemannian geometry presents an impressive array of special devices and procedures. Among these are coordinate systems (geodesic and Riemannian) having the property that at least one point $\left\{ {a \atop bc} \right\} = 0$, systems of N vector fields (orthogonal ennuples), operations based on mappings, and the use of anticommutative algebra (exterior differential forms).

Basic concepts. A relatively simple Riemannian geometry is furnished by a two-dimensional surface imbedded in ordinary euclidean space E_3. Specifically, let $\mathbf{r}$, the radius vector from a point O in E_3, be made a function $\mathbf{r}(x^1,x^2)$ of two real variables x^1, x^2 whose restrictions in size may be indicated by specifying that they are the rectangular cartesian coordinates of the points in a certain domain D of an auxiliary plane, E_2. Further let $\mathbf{r}(x^1,x^2)$ be single-valued and of class C^M, $M \geqq 1$, and meet the requirement $\mathbf{r}(a,b) \neq \mathbf{r}(c,d)$ for each pair of distinct points in D. This procedure defines a one-to-one reciprocal and continuous correspondence between the points in the two-dimensional domain D and the set R_2 of end-points of $\mathbf{r}(x^1,x^2)$. Thus, each admissible ordered number pair a, b ($x^1 = a$, $x^2 = b$) determines a point in D and simultaneously a point in R_2. Thus, the space R_2 bears a coordinate system, that is, a scheme for associating ordered number sets (pairs in the present case) to points in the space. The equations $x^2 = b$, with x^1 variable, and $x^1 = a$, with x^2 variable, determine a pair of coordinate lines in D and at the same time the parameterized arcs $\mathbf{r} = \mathbf{r}(x^1,b)$ and $\mathbf{r} = \mathbf{r}(a,x^2)$ (coordinate curves) in R_2. The derivatives $\mathbf{r}_a$ ($\mathbf{r}_1 = \partial\mathbf{r}/\partial x^1$, $\mathbf{r}_2 = \partial\mathbf{r}/\partial x^2$) at a point (a,b) in R_2 are tangent vectors to the two coordinate curves through (a,b). More generally, the equations $x^1 = x^1(t)$ and $x^2 = x^2(t)$ determine an arc C in D and also the mate arc $\mathbf{r} = \mathbf{r}[x^1(t), x^2(t)]$ in R_2. The vector $\mathbf{r}'$ ($\mathbf{r}' = D_t\mathbf{r} = \mathbf{r}_1 x'^1 + \mathbf{r}_2 x'^2 = \mathbf{r}_a x'^a$) is a tangent vector to C. The magnitude of $\mathbf{r}'$ expressed in surface quantities is given by

$$|\mathbf{r}'|^2 = [(\mathbf{r}_a x'^a)\cdot(\mathbf{r}_b x'^b)] = (g_{ab}x'^a x'^b),\ g_{ab} \overset{\text{def}}{=} \mathbf{r}_a\cdot\mathbf{r}_b \tag{1}$$

and Eq. (1) is valid for C considered as a curve in E_3 except that the range of indices is now 1 to 3 and x^1, x^2, x^3 must denote the variables of a coordinate system in E_3. For example, in terms of cartesian coordinates x, y, z, C is given by $\mathbf{r} = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{f}$, $\mathbf{r}_1 = \mathbf{i}$, $\mathbf{r}_2 = \mathbf{j}$, $\mathbf{r}_3 = \mathbf{f}$, $g_{ab} = \mathbf{r}_a\cdot\mathbf{r}_b = 0$ or 1 according to whether $a \neq b$ or $a = b$, and $|\mathbf{r}'|^2 = (x')^2 + (y')^2 + (z')^2$. Furthermore, according to a formula of calculus for Δs (arc length),

$$\Delta s = \int_{t_1}^{t_2} [(x')^2 + (y')^2 + (z')^2]^{1/2}\,dt$$

and so for C regarded either as a curve in R_2 or E_3 it follows that

$$\Delta s = \int_{t_1}^{t_2} |\mathbf{r}'|\,dt = \int_{t_1}^{t_2} (g_{ab}x'^a x'^b)^{1/2}\,dt \tag{2}$$

Thus, in summary R_2 is a space which bears a coordinate system (x) and there is associated with R_2 and (x) a set of functions g_{ab} which are involved in the determination of arc length. Obviously, the same may be said of E_3.

The quantities g_{ab} are involved in other metric relationships. For example, if C_1 and C_2 are two arcs in R_2 which intersect at a point P and if the values of x'^a for C_1 and C_2 are denoted respectively by U^a and V^a, then the angle between the tangents at P is given by

$$|\mathbf{r}'(C_1)|\,|\mathbf{r}'(C_2)| \cos A = \mathbf{r}'(C_1)\cdot\mathbf{r}'(C_2) = (\mathbf{r}_a U^a)\cdot(\mathbf{r}_b V^b) = g_{ab}U^a V^b \tag{3}$$

Also, it develops that a suitable extension of the concept area A from plane configurations to curved surface patches is furnished by the definition

$$A = \iint G^{1/2}\,dx^1\,dx^2 \quad G = g_{11}g_{22} - g_{12}g_{21}$$

For most of the developments in Riemannian geometry, it is necessary or convenient to impose the restriction that for a, b in D, the determinant $G(a,b) \neq 0$.

Because the quantities g_{ab} are functions of x^1 and x^2, they may be associated with the domain D and if other surfaces R_2 are introduced with the same domain of definition D, then there becomes associated with D more than one set of functions $g_{ab}(x)$, assuming that the same letters x^1, x^2 are used. Each set of g_{ab}s applied to mathematical entities in D determines measurements of related entities in the associated R. Thus, there arises a variety of problems concerning correspondences and mappings.

Generalizations to higher dimensions. The step from two to N dimensions can be made easily by making the coordinate variables $y^i(i{:}1$ to $L)$ of an L-dimensional euclidean space E_L functions of N variables x^a with $N < L$, thus $y^i = y^i(x^1, \ldots, x^N)$. If N of these equations (for example, the first N) determine the xs as functions of $y^1, \ldots, y^N$, $x^a = x^a(y^1, \ldots, y^N)$, then substitution into the remaining $L - N$ equations will relate the $y^i(i{:}N+1$ to $L)$ to $y^1, \ldots, y^N$. Thus, this procedure separates out a subset R_N of the points of the original E_L. If the coordinate variables y are cartesian and the symbols $\mathbf{i}_i$ denote the corresponding base vectors in E_L, then $\mathbf{r} = \mathbf{OP} = y^i(x)\mathbf{i}_i = \mathbf{r}(x^1, \ldots, x^N)$ is the vector equation of the subspace R_N. Examination of the development of Eqs. (1), (2), and (3) will show that these relationships may be taken over provided the range of indices is changed from 1, 2 to $1, \ldots, N$. Thus, the space R_N is Riemannian.

Transformation theory. Substitution from a set of transformation equations $x^a = x^a(\bar{x}^1, \ldots, \bar{x}^N) = x^a(\bar{x})$ of the general type of tensor analysis, into the vector function $\mathbf{r}(x)$ yields the mate function $\bar{\mathbf{r}}(\bar{x})$; thus, $\bar{\mathbf{r}}(\bar{x}) = \mathbf{r}[x(\bar{x})]$, and the equation of the subspace R_N in terms of the variables $\bar{x}$ is $\mathbf{r} = \bar{\mathbf{r}}(\bar{x})$. Differentiation of the identities $\bar{\mathbf{r}}(\bar{x}) = \mathbf{r}[x(\bar{x})]$, $\mathbf{r}(x) = \bar{\mathbf{r}}[\bar{x}(x)]$ yields the transformation equations

$$\frac{\partial \bar{\mathbf{r}}}{\partial \bar{x}^r} = \bar{\mathbf{r}}_r = \mathbf{r}_a X_r^a$$

$$\mathbf{r}_a = \bar{\mathbf{r}}_r X_a^r \left(X_r^a = \frac{\partial x^a}{\partial \bar{x}^r},\ X_a^r = \frac{\partial \bar{x}^r}{\partial x^a} \right) \tag{4}$$

It now follows from the definitions

$$\bar{g}_{rs} = \bar{\mathbf{r}}_r \cdot \bar{\mathbf{r}}_s\ g_{ab} = \mathbf{r}_a \cdot \mathbf{r}_b$$

and the relations

$$\bar{\mathbf{r}}_r \cdot \bar{\mathbf{r}}_s = (\mathbf{r}_a X_r^a)\cdot(\mathbf{r}_b X_s^b) = \mathbf{r}_a \cdot \mathbf{r}_b X_r^a X_s^b$$

that
$$\bar{g}_{rs} = g_{ab}X_r^a X_s^b \tag{5}$$

that is, the quantities g_{ab} and $\bar{g}_{rs}$ are tensor components of the type (0,2,0). An implication of Eq. (5) and the column-column rule for expressing the product of two Nth order determinants as an Nth-order determinant is if $\bar{G}$, G, and $J(x/\bar{x})$ denote the determinants of $\bar{g}_{rs}$, g_{ab}, X_r^a, respectively, then

$$\bar{G} = G[J(x/\bar{x})]^2 \tag{6}$$

It is convenient [illegible] a second set [illegible]

$\bar{g}^{rs}\bar{\mathbf{r}}_s$. The principal properties of the new base vectors are

$$\mathbf{r}^a \cdot \mathbf{r}_b = \delta_b^a\, g^{ab} = \mathbf{r}^a \cdot \mathbf{r}^b\ \mathbf{r}_a = g_{ab}\mathbf{r}^b \tag{7}$$

As a preliminary to the proof of these relations it may be noted that according to a formula of determinant theory $g_{bc}g^{ac} = \delta_b^a$. The symbol δ (called the Kronecker delta) may be defined by the statements $\delta_1^1 = \delta_2^2 = \cdots = \delta_N^N = 1$, $\delta_b^a = 0$ for $a \neq b$. It acts as a substitution operator, thus (for $N = 3$)

$$\delta_2^c\mathbf{r}_c = \delta_2^1\mathbf{r}_1 + \delta_2^2\mathbf{r}_2 + \delta_2^3\mathbf{r}_3 = \mathbf{r}_2$$

Also, $X_s^a X_b^s = \partial x^a/\partial \bar{x}^s\, \partial \bar{x}^s/\partial x^b = \partial x^a/\partial x^b = \delta_b^a$

since x^a and x^b belong to the same coordinate system. The Eqs. (7) may be established as follows: $\mathbf{r}^a \cdot \mathbf{r}_b = g^{ac}\mathbf{r}_c \cdot \mathbf{r}_b = g^{ac}g_{bc} = \delta_b^a$, $\mathbf{r}^a \cdot \mathbf{r}^b = g^{ac}\mathbf{r}_c \cdot \mathbf{r}^b = g^{ac}\delta_c^b = g^{ab}$, $g_{ab}\mathbf{r}^b = g_{ab}g^{bc}\mathbf{r}_c = \delta_a^c\mathbf{r}_c = \mathbf{r}_a$.

If a given vector α is expressible in the form $\alpha = a^a\mathbf{r}_a$, the quantities a^a are called the scalar components of α relative to the set $\mathbf{r}_a$. Since $\alpha \cdot \mathbf{r}^b = a^a\mathbf{r}_a \cdot \mathbf{r}^b = a^a\delta_a^b = a^b$, it follows that $\alpha = (\alpha \cdot \mathbf{r}^a)\mathbf{r}_a = \alpha \cdot (\mathbf{r}_b g^{ab})\mathbf{r}_a = (\alpha \cdot \mathbf{r}_b)\mathbf{r}^b$.

In particular,

$$\bar{\mathbf{r}}^r X_r^{\,a} = [(\bar{\mathbf{r}}^r X_r^{\,a}) \cdot \mathbf{r}_b]\mathbf{r}^b = (\bar{\mathbf{r}}^r X_r^{\,a}) \cdot (\bar{\mathbf{r}}_s X_b^{\,s})\mathbf{r}^b$$
$$= \delta_s^{\,r} X_r^{\,a} X_b^{\,s} \mathbf{r}^b = X_s^{\,a} X_b^{\,s} \mathbf{r}^b = \delta_b^{\,a} \mathbf{r}^b = \mathbf{r}^a$$

Also, $g^{ab} = \mathbf{r}^a \cdot \mathbf{r}^b = (\bar{\mathbf{r}}^r X_r^{\,a}) \cdot (\bar{\mathbf{r}}^s X_s^{\,b}) = \bar{g}^{rs} X_r^{\,a} X_s^{\,b}$

that is,

$$\mathbf{r}^a = \bar{\mathbf{r}}^r X_r^{\,a}\, g^{ab} = \bar{g}^{rs} X_r^{\,a} X_s^{\,b} \tag{8}$$

Thus, the quantities g^{ab} are tensor components of the type (2,0,0), that is, contravariant of order two.

The dual sets $\mathbf{r}_a$, $\mathbf{r}^a$ and g_{ab}, g^{ab} provide the basis for a dual representation of various entities. For example, if the contravariant components x'^a of the tangent vector $\mathbf{r}'(\mathbf{r}' = x'^a \mathbf{r}_a)$ for a given curve in an R_N in E_L are denoted by V^a and if $V_b \stackrel{\text{def}}{=} g_{ab}V^a$, then $\mathbf{r}' = V^a \mathbf{r}_a = V^a(g_{ab}\mathbf{r}^b) = V_b \mathbf{r}^b$.

Higher abstractions. A higher-order abstraction can be obtained by discarding the enveloping euclidean space E_L and considering the space R_N with a coordinate system (x) and a set of functions $g_{ab}(x)$ [illegible]

$|\alpha|^2 > 0$ and therefore the g_{ab}s are positive definite. The special and general theories of relativity involve forms $g_{ab}V^aV^b$ which can be positive, negative, or zero for real values of V^a.

In the present abstract case, certain of the preceding formulas may be adopted as definitions. Thus, a set of contravariant quantities V^a will be said to be the components of a vector and for brevity the set V^a *itself will be called a vector. The terms magnitude* of a vector and the angle A between two vectors $\mathbf{U}^a$ and $\mathbf{V}^a$ are defined by

$$|V^a|^2 \stackrel{\text{def}}{=} g_{ab}\mathbf{V}^a\mathbf{V}^b$$
$$\cos A \stackrel{\text{def}}{=} g_{ab}\mathbf{U}^a\mathbf{V}^b / \sqrt{(g_{cd}\mathbf{U}^c\mathbf{U}^d)(g_{ef}\mathbf{V}^e\mathbf{V}^f)}$$

The quantities $\bar{g}_{rs}$ associated with system $(\bar{x})$ may be defined as the coefficients of $\bar{\mathbf{U}}^r\bar{\mathbf{V}}^s$ in the transform $g_{ab}\bar{\mathbf{U}}^r X_r^{\,a} \bar{\mathbf{V}}^s X_s^{\,b}$ of $g_{ab}\mathbf{U}^a\mathbf{V}^b$. Thus, $\bar{g}_{rs} = g_{ab}X_r^{\,a}X_s^{\,b}$ and the quantities g_{ab}, $\bar{g}_{rs}$ are again tensor components of the type (0,2,0). The existence of the quantities g^{ab} requires that $G \neq 0$. In this case the covariant counterparts $\mathbf{V}_a$ of the contravariant $\mathbf{V}^a$ and $\mathbf{V}^a$ are related thus: $\mathbf{V}_a = g_{ab}\mathbf{V}^b$, $\mathbf{V}^a = g^{ab}\mathbf{V}_b$.

Equipollent displacement. Let $E(A)$ and $E(B)$ be the tangent planes at points A and B of a surface R_2 in E_3 with $E(A)$ not $\parallel$ to $E(B)$. Further, let α be a vector in $E(A)$, for example, at A with α not $\parallel$ to [illegible]

is equipollent to α relative to R_2 and preserve a substantial core of the cardinal properties of parallelism E_3. This can be done provided that the new concept be made relative to a curve C in R_2 joining A and B —*equipollence* (R_2, C).

Definition. Let C be an arc of class C'

$$\{x^a = x^a(t)\}$$

which lies in an $R_N(G \neq 0)$, and let $V^a(t)$ (class C') be a contravariant vector field along C. If $IV^a \underset{t}{=} 0$, then the field V^a will be said to be equipollent relative to R and C. Here I denotes intrinsic differentiation.

In particular, if A and B are points on C corresponding to the values t_1 and t_2 of the parameter t, then the vectors $V^a(t_1)$, $V^a(t_2)$ are equipollent (R_N, C). *If the quantities* $\left\{{a \atop bc}\right\} x'^c$ *are analytic, then the equations* $IV^a \stackrel{\text{def}}{=} V^{a\prime} + V^b \left\{{a \atop bc}\right\} x'^c = 0$, $V^a(t_1) = A^a$ will determine $V^a(t_2)$ uniquely. The Christoffel symbols $\left\{{a \atop bc}\right\}$ may be determined by $\left\{{a \atop bc}\right\} = \mathbf{r}^a \cdot \mathbf{r}_{bc}$ in the case of an R_N in E_L and more generally by

$$\left\{{a \atop bc}\right\} = g^{ad}\tfrac{1}{2}[-g_{bc,d} + g_{dc,b} + g_{bd,c}] \qquad (; \equiv \partial)$$

The notation $(; \equiv \partial)$ indicates that $;d \equiv \partial/\partial x^d$. Certain properties of intrinsic differentiation $[Ig_{ab} = Ig^{ab} = I\,\delta_b^{\,a} = 0,\ (IT_{ab})U^aV^b + T_{ab}(IU^a)V^b + T_{ab}U^a(IV^b) = (T_{ab}U^aV^b)'$ for T, U, V of types (0,2,0), (1,0,0), (1,0,0)$]$ ensure that the magnitudes of and the angles between vectors undergoing equipollent displacement remain constant. Also if E_a is the Euler vector for $F(x,x') = \sqrt{g_{ab}x'^ax'^b}$, $E_a \equiv -\partial F/\partial x^a + (\partial F/\partial x'^a)'$ and the parameter is the arc length s, then $I_s(dx^a/ds) = g^{ab}E_b$. An immediate consequence is that if C satisfies $E_b = 0$ (so that C is an extremal for $\int\sqrt{g_{ab}x'^ax'^b}\,ds$), then C satisfies $I_s(dx^a/ds) = 0$ (dx^a/ds is the unit tangent vector). The shortest arcs are extremal arcs; thus briefly stated the shortest arcs are the straightest.

Subspaces. If, in the generalization from two to N dimensions, the variables y^i are coordinates in an R_L (instead of E_L), then the subset of points determined by the variables x is an N-dimensional subspace [illegible] nian, [illegible] are

$\partial y^i/\partial x^a\, \partial y^j/\partial x^b\, x'^ax'^b = g_{ab}x'^ax'^b$, $g_{ab} \equiv g_{ij}\, \partial y^i/\partial x^a\, \partial y^j/\partial x^b$ The quantities $\partial y^i/\partial x^a$ for each value of i (1 to L) are covariant relative to transformations of the xs. Hence the g_{ab}s of R_N are determined by the g_{ij}s of R_L and are of type (0,2,0). In particular, if α and β are two perpendicular vectors at a point P of R_L, and $C(u)$ is the geodesic arc of R_L through P and tangent to $\cos u\alpha + \sin u\beta$, then the set of all such arcs is an $R_2[R_2(\alpha,\beta)]$ in R_L. The Gaussian curvature of $R_2(\alpha,\beta)$ is called the curvature $C(R_L,\alpha,\beta)$ of R_L for the orientation α, β. If $C(R_L,\alpha,\beta)$ is everywhere independent of the choice of α, β, then C does [illegible] ariant de- [illegible] is given [illegible] type (1,1,0).

Similarly, the covariant derivative of $V^a{}_{,c}$ with respect to x^d, namely, $V^a{}_{,c,d}$ is given by $V^a{}_{,c,d} = V^a{}_{,c;d} + V^e{}_{,c}\left\{{a \atop ed}\right\} - V^a{}_{,b}\left\{{b \atop cd}\right\}$ and $V^a{}_{,c,d} - V^a{}_{,d,c} = V^bR^a{}_{bcd}$ with $R^a{}_{bcd} = \left\{{a \atop bc}\right\}_{;d} - \left\{{a \atop bd}\right\}_{;c} + \left\{{e \atop bc}\right\}\left\{{a \atop de}\right\} - \left\{{e \atop bd}\right\}\left\{{a \atop ce}\right\}$. The quantities $R^a{}_{bcd}$ are the components of the celebrated Riemann-Christoffel tensor [type (1,3,0)]; some writers use $R^a{}_{bdc}$ for the $R^a{}_{bcd}$ given here.

Evidently, $R^a{}_{bcd} = 0$ is a necessary and sufficient condition for the equality of $V^a{}_{c,d}$ and $V^a{}_{d,c}$ with V^a of class C'', but otherwise arbitrary. Also, if for a certain coordinate system (x), the gs are constant (intrinsic flatness), then in $(x)\left\{{a \atop bc}\right\} \underset{a,b,c,x}{\equiv\!\equiv} 0$ and $R^a{}_{bcd} \underset{a,b,c,d,x}{\equiv\!\equiv} 0$. But $R^a{}_{bcd}$ is a tensor and $R^a{}_{bcd} \underset{a,b,c,d,x}{\equiv\!\equiv} 0$ implies $\bar{R}^r{}_{stu} \underset{r,s,t,u}{\equiv\!\equiv} 0$ in $(\bar{x})$. Consequently, if for some one coordinate system $(\bar{x})$ and some particular choice of r,s,t,u $\bar{R}^r{}_{stu} \neq 0$, then a coordinate system does not exist in the given R_V for which the gs are constant. The conditions $\left\{{a \atop bc}\right\}$ analytic, $R^a{}_{bcd} \equiv 0$ are sufficient for intrinsic flatness. For intrinsically flat R_V equipollence is independent of the arc.

The associated tensors R_{abcd} $(\equiv g_{ae}R^e{}_{bcd})$, R_{bd} (the Ricci-Einstein tensor, $R_{bd} \equiv R^a{}_{bad}$), and R (the scalar curvature, $R = g^{bd}R_{bd}$), are all of fundamental importance in Riemannian geometry and relativity. In particular, the Gaussian curvature of an R_2 is determined through the associated R_{1212}.

Some of the generalizations are: the geometry of paths (a nonmetric differential geometry based on quantities $\Gamma_{bc}{}^a$ which transform in the manner of Christoffel symbols and provide the definition of parallel displacement); Finsler geometry [a function $F(x,x')$ which is positively homogeneous of degree one in the x's takes the place of $\sqrt{g_{ab}x'^a x'^b}$]; higher order differential geometries (Kawaguchi spaces, F is a function of the xs, x's, and higher-order derivatives). *See* CALCULUS OF TENSORS; CALCULUS OF VECTORS; GEOMETRY, EUCLIDEAN; GEOMETRY, NONEUCLIDEAN.

[H.V.C.]

Geomorphology

The primary relief elements of the earth's surface are formed by movements of the crust (*see* DIAS-

(2) built up of the erosional debris. Most of the basic concepts of this branch of geology were formulated by William Morris Davis, of Harvard University, in the early part of this century. Davis held that the morphology of landforms is controlled by the interplay of three factors: structure, process, and stage.

Structure. In its Davisian sense, structure embraces all the characteristics of rocks which influence topographic form. The flat benches and vertical cliffs of the sides of the Grand Canyon are etched by erosion in layers of sandstone, shale, and limestone which are nearly horizontal. The sharply contrasted linear ridges and valleys of the Appalachians correspond in position with the edges of folded strata. Because most erosional processes are in the highest degree selective, differences in permeability, solubility, and other properties of rocks, so subtle as to be difficult to measure in the laboratory, are commonly reflected in surface form. The geomorphologist looks first for structural causes of landform anomalies before considering other explanations.

Process. The processes of erosion differ markedly from place to place depending on climate and other circumstances, and each process tends to develop its characteristic landforms. A cirque is a distinctive feature, carved by the head of a valley glacier. A wave-cut cliff is similarly unique in form and uniquely associated with a specific erosional process. In the absence of a vegetative cover, gully washing may cut an area into a plexus of badland channels and knife-edged divides wholly different from the smoothly contoured landscape with broadly convex hill tops that is modeled in the same material by downhill creep of plant-held soil. A knowledge of the process, preferably by direct observation, contributes much to an understanding of the form, and the reverse is equally true.

The relationship between process and form, once established, becomes the basis for a rich variety of interpretations. Cirques prove the former existence of glaciers where there are none now. A wave-cut cliff high above the modern strand line proves a relative change in level of land and sea. A smoothly modeled hillslope cut by gullies is a compound landform, indicating a change in process. Most present-day landforms are compound, with relict elements formed by Pleistocene processes being modified or destroyed by processes suited to present conditions.

Stage. Stage expresses change in landforms with the passage of time. If the relative level of land and sea remains the same long enough, the Colorado River will eventually reduce its gradient to just that slope which will enable the river to transport to the sea all of the erosional debris shed into it from its valley sides. Continuous loss of material means that the valley sides will be gradually reduced in slope and height until the Colorado Plateau comes to be a peneplain, that is, an erosional lowland of faint relief, drained by sluggish rivers meandering on broad valley floors. This is the concept of the geomorphic cycle. The progress of the cycle cannot be seen in any one place because the changes are exceedingly slow, but over the face of the earth, assemblages of landforms representing every phase of it can be observed. The stage of maturity is characterized by an integration of all parts of the area under consideration in a system of slopes graded to the sea. Disadjustments of various types, as lakes and waterfalls that interrupt the profiles of the streams, and analogous breaks in slope in the interstream areas, are symptoms of "youth." The term "old" suggests the greatly reduced vigor of the erosional processes as the peneplain is approached.

Validity. There has been debate as to many of the details of the Davis geomorphic cycle, but there is no question as to the validity of its central theme: that during a period of crustal stability, landforms develop in an orderly and predictable manner with respect to sea level. Actually, sea level oscillated markedly during Pleistocene time under the influence of glacioeustatism, and the

crust has been exceptionally unstable in late Cenozoic time. Far from reducing the value of the cycle concept, these circumstances make it indispensable, because they mean that most assemblages of landforms include elements of two or more cycles that would, in most cases, not be distinguished without the insight provided by this concept. *See* FLUVIAL EROSION CYCLE.

The geomorphic cycle as outlined above applies primarily to the geomorphic features of the humid temperate zone where the science had its birth. It differs in detail and degree rather than in kind when applied to the sequences of landforms developed under polar and tropical conditions, and in arid regions where there is no drainage to the sea. The same basic scheme has proved its worth in analysis of the wholly different kinds of geomorphic features formed by shoreline processes, glaciers, ground water, and other agents of erosion.

Every erosional landform expresses in its shape the influence of structure, process, and stage. Because each of these factors is complex, it is understandable that erosional landforms are almost infinitely varied. Depositional landforms—deltas, glacial moraines, offshore bars, sand dunes and other forms—are equally varied in details of morphology. This helps to explain why it is that until recently the approach to geomorphic problems has been chiefly qualitative. Since about 1950 the most significant development has been an increase in the application of quantitative and experimental methods in geomorphic research. These methods will do much to clarify and refine the understanding of geomorphic processes and forms. *See* COASTAL LANDFORMS; DESERT EROSION FEATURES; FLUVIAL EROSION LANDFORMS; GLACIATED TERRANE; KARST TOPOGRAPHY. [J.H.M.]

Bibliography: N. E. A. Hinds, *Geomorphology*, 1943; A. K. Lobeck, *Geomorphology*, 1939; W. D. Thornbury, *Principles of Geomorphology*, 1954; O. D. von Engeln, *Geomorphology*, 1942.

Geophysical exploration

The investigation of earth character and the search for natural resource deposits by measurement of physical properties of an earth feature such as an ore deposit; examples are the electrical conductivity, density, or radioactivity of an ore body. Some features are studied less directly by measuring the properties of the confining structures, for instance, anticlines, salt domes, or fault traps in petroleum prospecting. Techniques are available to measure (1) naturally occurring phenomena (gravity, electrical and magnetic fields, for example) or (2) induced effects, such as seismic waves resulting from planned explosions or electric currents generated in the earth by sources of electrical energy. The measurements are commonly made on the surface of the earth, in aircraft, on the surface or bottoms of inland or coastal waters, or in bore holes which may be available or expressly drilled for that purpose.

Intensive development and application of geophysical exploration techniques began shortly after the close of World War I, predominantly in the United States. Perfection of these techniques has been largely dependent on the development of instruments both highly portable and capable of sensitivities (under field operating conditions) previously found exclusively in laboratory environments. For example, modern gravity meters permit the detection of gravity differences of 1 part in 100,000,000.

The required observations usually are made at regular intervals along a line or at intersections of grid lines laid over a map of the area to be investigated. The map, drawn accurately to scale, also shows the bench marks on the ground. These marks are necessary so that any point on the map can be identified on the ground. This is the base map for compiling the results of field observations. Commonly lines are drawn connecting points of equal field measurements; these are similar to the contour lines of equal elevation on topographic maps. Departures from uniformity on such maps are known as anomalies, and establishment of their geological significance is the objective of geophysical interpretation.

The surface or weathered layer (such as soil or subsoil) often distorts geophysical measurements in amounts disproportionate to its thickness. This is especially true of areas covered by glacial drift or where gravel beds lie at or near the surface. In planning observational procedures careful consideration must be given to possible effects of the surface layer if sound results are to be obtained.

Geophysical techniques also are used effectively in other fields of activity. This is particularly true in foundation planning for roads, dams, bridges, and buildings and also in archaeological exploration.

Gravity exploration. Variation in gravitational acceleration at any point on the earth, although largely determined by the mass, size, and shape of the entire earth, depends appreciably on the rock masses in the immediate environment of the point. Because of the earth's spheroidal shape (slight polar flattening and equatorial bulge), gravity decreases from the equator to the poles (latitude variation). It varies also with elevation at any latitude, as well as with the topography in the environment of the observation point. Observed gravity values are adjusted for these and other effects. The residual values are presumed to be related directly to the local geology. This is the basis of gravity exploration *See* TERRESTRIAL GRAVITATION.

Gravity surveys are used more extensively for petroleum exploration than for metallic mineral prospecting. It is true that the density difference between metallic ores and surrounding earth is relatively greater than that between oil deposits and surrounding earth; but the size of ore bodies is generally moderate, and therefore the mass excess, the difference between the ore body mass

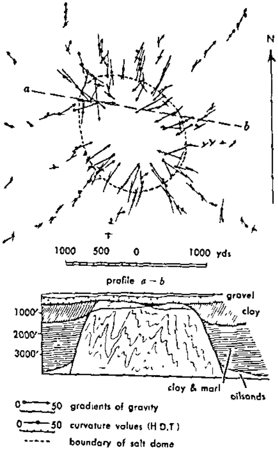


Fig. 1. Diagram illustrating gravity gradients and HDT (horizontal directive tendency) over a subterranean salt dome. Small arrows at top show direction and strength of gravity gradients at different points. (After A. S. Eve and D. A. Keys, Applied Geophysics, Cambridge, 1954)

the mass of an equal volume of the rock surrounding it, is rather small. The gravity effects of ore bodies are usually quite local and small, and relevant gravity surveys need to be detailed. The deeper the ore bodies lie, the larger they have to be to yield an observable gravity effect. In petroleum prospecting, on the other hand, the density differentials are smaller in order of magnitude, but the generally greater dimensions more than offset this. Thus, mass excesses or deficits are substantial.

Almost all gravity measurements are relative ones; differences between observation points are determined, and their absolute values remain unknown. An arbitrary value is assigned to a base point, and all other values are relative to it. The distance between stations should not be greater than one-half the depth to the structures being studied by d... as small as ½ mile in between te...ical needs and cost is required. In addition to the gravity value, the position and elevation of the observation point must be rather precisely determined.

Instrumentation. The instrument most used in gravity surveys is the gravity meter (gravimeter), which has been developed in many forms. Its important characteristics are sensitivity, scale linearity, stability, low-level sensitivity, low-temperature response, minimal drift, and of course, portability. There are special designs for use under water. Quite recently, stabilized platforms have been developed which may make gravity surveys on surface ships possible.

In its simplest form, a gravimeter consists of a mass suspended from a spring. An increase of gravity increases the weight of the mass and causes a corresponding extension of the spring. Gravimeters are calibrated by observing the deflection resulting from the addition of a known weight to the suspended mass or by measuring at two stations for which the gravity difference has been otherwise determined.

In the early days of gravity exploration, the Eotvos torsion balance was used extensively. It measured the horizontal gradient of gravity over a horizontal distance of only about 40 cm. Its extreme sensitivity to local erratics and the hours required for an observation made the torsion balance obsolete as soon as gravimeters became available. The gravity pendulum, occupying an intermediate stage in prospecting, also yielded to the gravimeter, which excelled it in a number of respects, especially unit costs.

Cartographic technique. Gravity maps display highs and lows (hills and valleys). These are referred to as regional if they cover some hundreds or more square miles, subregional if their area is some tens of square miles, and local if they extend only over an area of some square miles or less. Usually, it is the local anomalies which are of direct interest in natural resource exploration. The nature of these local anomalies is determined in part by the depth to the mass anomaly with which they are related, and an estimate of this depth can usually be made from the gravity data alone.

Rapid computational procedures are available for computing the gravity anomaly corresponding to an assumed structure and density distribution. Experience and the character of the anomaly will suggest appropriate assumptions. The computed anomaly then can be compared to the observed one. In several tries, the assumed mass anomaly usually can be made to account for the observed gravity anomaly. Unfortunately, a unique answer cannot be obtained; but a knowledge of the geology general and specific to the area can reduce the uncertainty to within acceptable limits. A less sophisticated approach avoids the calculations. A test hole is drilled at the anomaly site. A depth computation would indicate how deep the hole must be to reach the anomaly source.

Magnetic exploration. Magnetic techniques are based on the fact that the geomagnetic field is locally modified by the magnetization of surface and near-surface rocks. Under certain circumstances, this modification can be substantial. Mag-

netite, Fe_3O_4, is by far the most magnetic common mineral. Its presence accounts for the magnetization of most igneous and sedimentary rocks. Instances are known in which the magnetic effect of magnetite existing in massive form (iron formation) exceeds the normal magnetic field. In igneous terrains, anomalies of 10% of the normal field intensity are not uncommon. In contrast, a local gravity anomaly is exceptionally large if it amounts to 0.001% of the total gravity field. *See* GEOMAGNETISM.

Igneous rocks are permanently magnetized when they cool through the Curie point of magnetic susceptibility. The direction and intensity of their permanent magnetization will depend on the direction and intensity of the then existing geomagnetic field. In addition, the rocks are magnetized by induction, the direction and intensity depending on the differences between original conditions and the geomagnetic field in which they now exist. Both the magnitude and direction of the geomagnetic field change slowly with time; furthermore, rocks change their position and orientation when involved in orogenic movements. Therefore, directions of the permanent and induced magnetizations may not be the same. The magnetic susceptibility of sedimentary rocks is generally less than that of igneous rocks by several orders of magnitude. Consequently, it is reasonably safe to assume that the magnetic anomalies observed in magnetic surveys of sedimentary basin areas arise from topographic or magnetization effects at the surface of or within the igneous basement.

Magnetite excepted, metallic minerals that constitute ore bodies are only slightly magnetic. Nevertheless, important ore bodies have been discovered as a result of magnetic surveys. Sometimes the association between the ore sought and magnetism is indirect. In placer deposits, gold often is concentrated in stream channels along with magnetite, and finding the magnetite concentrations may lead to the discovery of gold.

Within or near population centers, power and transportation systems may cause erratic fluctuations of the magnetic field. Under such conditions, it is often impossible to adjust instruments or make meaningful field measurements. There is also a diurnal cyclical variation of the magnetic field for which field measurements must be adjusted in areas where the magnetic relief does not exceed several hundred gammas. In areas where the relief is many hundreds or thousands of gammas, this variation is often ignored. Occasionally, there are variations much more rapid and greater than the diurnal ones. These are known as magnetic storms. Surveys made during them may be of inferior quality or even useless.

The magnetic measurements made for exploration purposes are relative ones. Modern techniques of interpretation require that they be made at points in a uniform grid. When such a grid involves serious operational difficulties, the uniform grid may be derived from an irregular one by interpolation. In ground measurements, the vertical [illegible]

sentially of a magnet system with a horizontal axis in which the turning moment arising from the magnetic field is compensated by the gravity moment resulting from an off-center weight. Compensating magnets also are used to compensate partially the field intensity, thus extending the measuring range of the instrument. This compensation is alternatively accomplished by varying the current in a Helmholtz coil placed around the suspended magnet system of the balance. Similar instruments are available for measuring the horizontal component. Finally, there are field instruments which serve to measure all the elements necessary to determine the total intensity and its direction in space. These so-called universal magnetometers have been found cumbersome to use and as a result are not widely used.

Immediately after World War II, an airborne magnetometer became available. The field-sensitive element of this is a rodlike assemblage of μ metal (mu metal, or other high initial-permeability material) with suitable coil windings and mounted in mutually perpendicular gimbal systems. A servomechanism maintains the μ-metal axis automatically in the direction of the total magnetic intensity. The variations of the total intensity are recorded on a roll of paper which advances under the recording pen at a uniform time rate. The corresponding position information is obtained by photographing the ground vertically below or by Shoran or other radiolocation device. Auxiliary means provide the necessary coordination between the magnetic and position data. The survey aircraft flies parallel lines at planned distance intervals and at a fixed barometric altitude, previously determined to be most suitable to terrain and interpretation requirements. The magnetic information is continuous along flight lines. Aeromagnetic surveys are mass-production operations. The cost of initiating them is quite high, and the unit cost factor favors surveys of large areas.

During the 1950s, an entirely new type of magnetometer was developed. In it, the measurement depends on the magnetic properties of protons. In a commercially available form, the protons required are provided by a mass of water, and the instrument is referred to colloquially as the bucket-of-water magnetometer. The measurement [illegible]

being measured. The precession frequency is a linear function of this field intensity. Since the precession is a transient phenomenon, the measurement must be made within a second or two. Electronic gear is used for making the frequency determination. The great advantage of this instrument is that its measurements are absolute.

Electrical exploration. The literature of these methods, which are used primarily in metallic mineral prospecting, is quite extensive, and their theory has been elaborated rather fully. Some mineral deposits give rise to spontaneous earth currents in their immediate environment, and the attendant electrical potentials are known as self-potentials. By mapping equipotential lines in any area, the source of the self-potential may often be found. There are also telluric currents in the earth that affect larger areas and are believed to be related to current circulation in the upper atmosphere. These also contribute to the local distribution of natural electrical potentials. The distribution of telluric currents has been studied to a limited extent in an endeavor to elucidate regional geologic structure. *See* TERRESTRIAL ELECTRICITY

The electrical methods most used generate artificial currents in the earth, either by conduction or by induction. A key problem is the provision of equipment which will generate sufficient electrical or electromagnetic energy in the ground and which will also be reasonably portable. The depth of current penetration depends on the geometry of the disposition of instruments, on the frequency used, and on the conductivity from the surface down. There is a distinct advantage in making measurements at several frequencies with the same instrument set-up, but this is possible only at the expense of portability.

The electromagnetic methods now are preferred greatly by the mining industry. They generally involve a transmitting coil, which is excited at a suitable frequency, and a receiving coil, which measures one or more elements of the electromagnetic field at a number of observation points. The receiving coil is usually so oriented as to minimize its direct coupling to the transmitter, and the residual effects are then due to the currents which have been induced in the ground. The results may reveal conductivity anomalies, some of which may represent ore bodies.

Electrical methods have become airborne. The transmitting and receiving coils and all associated gear are carried in an aircraft which normally flies as close to the ground as is consistent with safety.

Fig. 2. Graph showing use of depth resistivity curve (from single electrode probe) in determining depth of rock layers for a prospective dam site (*After A. S. Eve and D. A. Keys, Applied Geophysics, Cambridge, 1954*)

Electrical methods also have found effective use in exploring for ground-water accumulations in observing rock structures at dam sites. In the form of well-logging procedures, they are used a great deal in petroleum exploration and exploitation. They are used also for detecting the position of buried pipe lines, in land-mine detection, and in other military operations.

Seismic exploration. Fundamentally, this involves the initiation of elastic waves in the earth by explosions or other means of excitation, and interception of these waves by geophones (transducers, seismometers), which convert them into electrical pulses which are recorded on the uniformly moving tape of an oscillograph. The instant of wave initiation (shot moment) is transmitted to the recording gear by wire or radio signal. The side by side traces on the tape (one for each geophone) show the zero time for the waves and the complex of wave trains, which reach each geophone at times depending on the wave type and wave path. Association of the wave trains with their types and paths is facilitated by arranging a number of geophones in regular patterns, for instance, equispaced on a straight line which passes through the shot point. The geophones used chiefly respond to the vertical component of the earth motion. *See* MODEL THEORY.

The wave trains identifiable on the records consist largely of reflected and refracted longitudinal waves and of interface waves which have both longitudinal and transverse components. The latter include the surface waves colloquially referred to as ground roll.

The records serve to determine, to the nearest thousandth of a second, the travel time of any wave train from the shot point to the seismometer. The task of the interpreter is to infer from this time, and from general principles, the geometry of the path traversed by each wave train and, consequently, the position and attitude (slope) with respect to the operational surface of the surfaces which reflected or refracted the waves.

Refraction technique. Two broad categories of seismic methods are used in prospecting. They are the refraction and reflection techniques. To gain an insight into refraction shooting, it is assumed that a shot point and six or more equispaced geophones have been arranged along a straight line. The shot having been fired and the record made, each trace gives the travel time from the shot point to one geophone position for the first arriving wave. On graph paper, one plots these times against the corresponding shot point to geophone distances.

Figure 3*a* shows a typical time-distance graph. The geophone distance interval Δd divided by the corresponding time interval Δt represents the velocity between the two points. This ratio is also the reciprocal of the slope of the lines AB and BC. The segment AB therefore defines a velocity v_1 which is uniform because AB is a straight line, and the segment BC defines a higher uniform velocity v_2.

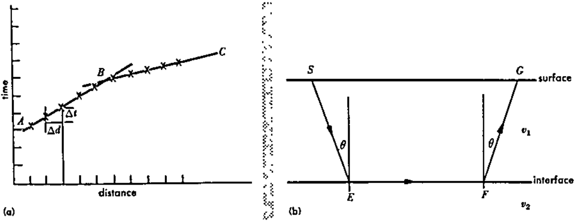


Fig. 3. Graphic plots of time and distance data recorded on geophones from a single shot. (a) Time-distance graph. (b) Graphic diagram of minimum time path.

It follows, with time-distance results like those shown, that below the surface on which the shooting was done, there lies an interface above which the velocity is uniformly v_1 and below which the velocity is v_2. For all points on AB, the waves first arriving have travelled on a direct line from shot point to geophone. For all points along BC, the wave path has been one such as $SEFG$ in Fig. 3b, the angles of incidence and emergence at the lower medium being given by $\sin \theta\ v_1/v_2$ The distance corresponding to the intersection of AB and BC in Fig. 3a depends only on the two velocities and the depth to the interface and is known as the critical distance. From it and the two velocities, the depth to the v_2 layer in Fig. 3b is readily computed.

A great advantage of the refraction method is that it provides information about the characteristic velocity of the refracting medium as well as its depth. The reflection method furnishes depth values only.

Refraction shooting is effective for high-speed layers at considerable depth, although the shot hole-seismometer distances become a matter of miles and the amounts of explosives needed become huge. Circular and fanlike seismometer arrays have been used as well as linear ones. Fan shooting has contributed materially to salt dome discoveries The seismic wave speed in salt is nearly always substantially greater than it is in the sediments surrounding the domes. Under such circumstances, the travel time for a given distance would be less than normal for any wave passing in part through massive salt.

Reflection technique. The reflection technique is the one now most extensively used in prospecting. It is essentially an echo-sounding procedure. Energy is reflected at each discontinuity where the velocity changes by virtue of changes in elasticity, density, or both. Such changes occur at geological formation boundaries. In general, the seismometers are set up quite near to the shot point—within 1000–2000 ft. The distance to the seismometer farthest out will be short if the reflection level in which the operator is interested is shallow, and longer as the depth increases. Usually the operator is primarily interested in reflections from a specific depth level. He will know approximately at what times these should appear on the record and will arrange his seismometer array so that unwanted waves will not arrive near this time.

What is measured is the total time interval required for the waves to go down to the reflection surface and to echo back. This time will increase with distance between the seismometer and the shot point. The reflection pulses, therefore, will angle across the record, the angle depending on the depth to and the slope of the reflecting horizon. Each record will provide time evidence on the depth and slope of one or more reflecting horizons. If any reflection can be confidently correlated between records, a time contour map of the reflecting horizon can be drawn. In some areas, such correlation is difficult or impossible In such cases, the slopes of reflections recorded during a limited time can be used to map a phantom horizon which is consistent with all the dips.

In order to convert the observed times into distances, the velocity must be known. Reflection observations do not yield this information directly. Experienced operators can usually make quite good estimates of the velocity Any error in this estimate will accentuate or minimize the subsurface topography and over- or underestimate the depth. Alternatively, appropriate shooting may be done at a well site which, along with the well log, provides the necessary velocity information.

The raw data produced by seismic methods are time intervals of some seconds, with differences measured in milliseconds, all related to the geometry of the shot-seismometer spread and to surface and subsurface geology. The practical success of the operation will depend critically upon how accurately the interpreter can visualize the wave paths which account for these time intervals. In oil prospecting, where a layered sedimentary section is usually involved, this is far easier than in the generally complex igneous terrains enco

the exploration for metallic minerals. *See* SEISMOLOGY.

Geochemical exploration. This is based on the concept that, through various natural processes, the surface soil and water environment of a buried natural resource deposit might contain minor quantities of chemical compounds derived from the deposit. Examples of such processes are seepage through pores or fissures, fluctuations in the level of ground water, and diffusion. It would be expected that the concentration of such diagnostic substances would increase toward the source and reach a maximum in its immediate vicinity. Giving due consideration to factors which may distort such a pattern, use of this concept in prospecting has met with some success.

In petroleum exploration, extensive experiments have been made involving chemical analyses of both soil gases and soils. At best the results of these experiments have been controversial. In metallic mineral exploration the situation is much more favorable. Here progress has been greatly facilitated by the development of many simple procedures for determining the presence of specific metallic elements, or groups of them, and closely estimating their amounts when present in concentrations measured in parts per million. Often these tests can be made in the field where the samples are being collected. Alternatively, the soil, water, or plant samples involved are sent to laboratories specially equipped for making such determinations in large numbers.

The use of plant material in such analyses is referred to as *biogeochemical* and also as geobotanical prospecting. Various species of plants, shrubs, and trees differ selectively in their take-up of elements from the soil. Sampling is preferably confined to a single species prevalent throughout the survey area. *See* GEOCHEMICAL PROSPECTING.

Radioactivity exploration. The detection and measurement of radiation emitted by radioactive elements such as uranium, thorium, and their disintegration products form the bases for such techniques. These radiations are completely absorbed within a few feet of soil cover and rapidly attenuated even in air. The Geiger counter is the instrument generally used to detect and measure the radiation. It is used mostly on the ground but also is used in bore holes and in low-flying aircraft. During the uranium rush of the 1950s, it could be bought in many local stores in the states most heavily prospected. Although radioactivity is chiefly used in the search for uranium ores, there are others, such as columbium ores, which may contain subsidiary quantities of uranium and as a consequence are moderately radioactive. For the use of radioactivity methods in oil exploration *see* PROSPECTING; WELL LOGGING (MINERAL).

[E.A.E.]

Bibliography: M. B. Dobrin, *Introduction to Geophysical Prospecting*, 2d ed., 1960.

Geophysics

Those branches of earth science in which the principles and practices of physics are used to study the earth. Geophysics is considered by some to be a branch of geology, by others to be of equal rank. It is distinguished from the other earth sciences largely by its use of instruments to make direct or indirect measurements of the parts of the earth being studied, in contrast to the more direct observations which are typical of geology. *See* GEOLOGY.

SUBDIVISIONS OF GEOPHYSICAL SCIENCE

Geophysics consists of several principal fields plus parallel and subsidiary divisions. These are commonly considered to include plutology, with geodesy, geothermometry, seismology, and tectonophysics as subdivisions; hydrospheric studies, mostly hydrology (ground-water studies) and oceanography; atmospheric studies, traditionally meteorology and recently aeronomy; and several fields of geophysics which overlap one another, including geomagnetism and geoelectricity, geochronology, geocosmogony, and geophysical exploration and prospecting.

Plutology. This is a general term covering geophysical methods of studying the solid part of the earth. The field has the following four major subdivisions.

Geodesy. The science of the shape and size of the earth is geodesy. This includes consideration of the earth's mass distribution as determined by measurements of the gravitational-centrifugal field (*see* GEODESY; TERRESTRIAL GRAVITATION). Because of man's interest in space travel, geodesy is concerned also with the distribution of the gravitational field above the earth's surface out to the limit of the earth's detectable effect.

Geothermometry. This is the science of the earth's heat (*see* GEOLOGIC THERMOMETRY; HIGH-PRESSURE PHENOMENA). This includes study of temperature variation and of heat generation, conduction, and loss in the earth, and their effects on the materials of which the earth is composed. Volcanology, the science of volcanoes, uses some of the principles and techniques of geothermometry.

Seismology. The science of earthquakes and other ground vibrations is the field of seismology (*see* SEISMOGRAPH; SEISMOLOGY). Seismology has made particularly important contributions to man's knowledge of the earth's interior. Study of the times of passage of seismic waves through the earth gives information on the distribution of different types of rock. *See* EARTH INTERIOR.

Tectonophysics. Sometimes called geodynamics, this is the science of the deformation of rocks. It consists of tectonics, the study of the broader structural features of the earth and their causes, as in mountain building; and rock mechanics, the measurement of the strength and related physical properties of rocks. *See* OROGENY; ROCK MECHANICS; TECTONOPHYSICS.

Hydrospheric studies. Geophysical study of the hydrosphere has two main branches, hydrology and oceanography.

Hydrology. This ground-water science also includes glaciology, the study of ground water in the form of snow and ice. *See* GLACIER; GROUND WATER; HYDROLOGY.

Oceanography. The scientific study of the oceans includes the study of the shape and structure of the ocean basins; the physical and chemical properties of sea water; ocean currents, waves and tides; thermodynamics of the oceans; and the relation of these to the organisms which live in the sea. *See* OCEANOGRAPHY; OCEANS AND SEAS; SUBMARINE TOPOGRAPHY.

Atmospheric studies. Because of extending interest from the face of the earth toward outer space, aeronomy has come to be recognized as a distinct science separate from meteorology.

Meteorology. The science of the earth's atmosphere yields meteorological information not only used in climatology and weather prediction, but also important in understanding the problems of aircraft flight and air pollution.

Aeronomy. This field is concerned with the phenomena of the upper atmosphere above 100 km, where its electrical behavior is important because the air is strongly ionized. There, where the mean free path of an atom or electron is long, the physical behavior of the material is controlled at least as much by its electrical properties as by its density and other mass properties. The outer boundary of the atmosphere is not a distinct surface; there is instead a gradual transition to the relative emptiness of interplanetary space. In this transition zone there is continuous interaction between the matter of the earth's atmosphere and the radiations and particles arriving from outside. Aeronomy is much concerned with these phenomena and their significance to earth conditions. Understanding of them is essential for the safe flight of man beyond the lower reaches of the earth's atmosphere. *See* AERONOMY; ATMOSPHERE; IONOSPHERE; METEOROLOGY.

Overlapping fields of geophysics. In addition to the regional subdivisions, there are several other fields of geophysics which overlap and concern all the others.

Geomagnetism, geoelectricity. The science of the earth's magnetic field and magnetic properties is termed geomagnetism. Geoelectricity is the science of electric currents in the earth and of its electrical properties. Geomagnetism and geoelectricity are closely related because the magnetic field is due largely to electrical currents in the solid earth and in the atmosphere. *See* GEOMAGNETISM; TERRESTRIAL ELECTRICITY.

Geochronology. This field deals with the dating of events in the earth's history. The principal technique used is based on radioactive disintegration. The proportions of parent to daughter elements in a mineral or rock are a measure of the age of the material. Other methods depend on the red shift of the spectrum of distant stars, the rate of recession of the moon, and the rates of erosion and sedimentation. *See* DATING METHODS; GEOCHRONOMETRY.

Geocosmogony. This is the study of the origin of the earth. The many hypotheses proposed fall into two groups, those which postulate that the earth is primarily an aggregate of once smaller particles, and those which claim that it is largely a fragment of a larger body. Current speculation favors the former theory. Geocosmogony is intimately linked with the origin of the solar system and our galaxy. Many lines of evidence suggest that the formation of the earth was a typical minor event in the evolution of the Milky Way or of the universe as a whole, occurring $5\text{–}6 \times 10^9$ years ago. *See* COSMOCHEMISTRY; COSMOGONY; GALAXY, THE.

Exploration and prospecting. Geophysical techniques are widely used not only to study the general structure of the earth, but also in prospecting for petroleum, mineral deposits, and ground water, and in mapping the sites of highways, dams, and other structures. Seismic methods are the most widely used, but electrical, electromagnetic, gravity, magnetic, and radioactivity surveying methods are also well developed. Many types of geophysical surveys can be made by lowering measuring apparatus into bore holes. *See* GEOPHYSICAL EXPLORATION; PROSPECTING; PROSPECTING, PETROLEUM.

Technical literature. Some of the principal geophysical publications largely in English are the *Journal of Geophysical Research* of the American Geophysical Union; *Geophysics*, published by the Society of Exploration Geophysicists; the *Bulletin of the Seismological Society of America;* the *Geophysical Journal of the Royal Astronomical Society;* the *Journal of Meteorology;* the *Bulletin of the American Meteorological Society;* the *Quarterly Journal of the Royal Meteorological Society;* the *Journal of Atmospheric and Terrestrial Physics;* and the *Bulletin of the Earthquake Research Institute of Tokyo.* Geophysical papers are commonly found also in the *Bulletin of the Geological Society of America; Geochemica and Cosmochemica Acta*, the journal of the Geochemical Society; the *Transactions of the American Institute of Mining, Metallurgical, and Petroleum Engineers;* and the *Proceedings of the National Academy of Sciences of the United States of America.* [B.F.HO.]

PROGRAMS OF GEOPHYSICS

Geophysical studies of the earth require data gathered from all parts of the globe for optimum success in delineating the structure of this planet and understanding the processes modifying it. For this reason the nations of the world cooperate in the exchange of data through various international agencies and scientific societies such as the I national Union of Geodesy and Geophysi

United States chapter is the American Geophysical Union. Periodically, as during the International Geophysical Year (IGY) in 1957–1958, a special effort is made to gather data systematically on a world-wide basis.

Development of scientific research has seldom seen more activity in a relatively short time than during the IGY. This program, or series of programs, has given the science of geophysics a much wider significance than it had prior to the IGY. In part this reflects conditions ripe for a synthesis of the geophysical sciences and for their extension. In some scientific fields observations had been made for decades, and in others for centuries. Particularly since 1935 great advances in electronic instrumentation have made possible accurate measurements of value in geophysical investigations. Special tools such as rockets and satellites vastly extend the range and scope of possible experiments. The advent of electronic computers and automatic data-processing devices makes possible the handling of monumental bodies of data that characterize some of the fields of geophysics—weather and ionospheric physics, for example. Thus, the IGY was successful in advancing geophysical studies because of advances in related sciences as well as from care and coordination in planning.

Solar-planetary aspects. The role of the sun in affecting terrestrial phenomena has long been recognized. It was largely, however, the fact that the IGY was carefully planned to coincide with a period of high solar activity, combined with the varied and extensive solar observations which were carried out during the IGY, that emphasized solar physics as a part of geophysics as well as of astronomy. Future geophysicists, particularly those who study the problems of oceans and air, will probably include in their programs the study of the sun. Another notable example of this extension of geophysics is the study of other planets. Many problems of interest to the geophysicist are aided by a study of related problems in other planets. The possibilities for increased accessibility of other members of the solar system have initiated a general use of the term department of planetary physics.

Satellite and rocket tools. Perhaps the most effective new tools available to the geophysicist are satellites and rockets. These tools give direct access to regions of the earth and its surrounding space that can be obtained in no other known way. The impact of the earth satellite on geophysics appears to be considerably greater than that of the rocket. Rockets have made great contributions to the study of the high atmosphere and particularly to the nondirect studies of the sun. In addition, they have led to the extension of rocket studies to satellites.

The contribution of satellites and moon rockets to an understanding of charged particles of considerable energy is effectively demonstrated by the discovery of the Van Allen radiation belts, an outstanding discovery during the IGY (see VAN ALLEN RADIATION). The theoretical, observational, and laboratory problems developed by the increased accessibility provided by satellites are so large in number, so difficult and exacting, that their importance can scarcely be overemphasized. With this one tool alone, certain areas of geophysics have taken on new stature and an attraction for an expanding group of scientists.

In addition to fields which will be strongly affected by satellite capabilities, such as ionospheric physics, geodesy (gravity), cosmic rays, and geomagnetism, is the large and theoretically challenging field of meteorology. A forward look indicates that by 1985 meteorology will be one of the most intellectually stimulating areas of investigation in geophysics.

Coordinating and planning. Plans are both intensive and extensive. The list of technical panels and committees responsible for planning and carrying out the United States program of the IGY shows not only the rapidly increasing breadth of the geophysical sciences, but also the intensive extension of geophysical observations to parts of the earth hitherto only relatively casually observed. The large Antarctic scientific program carried out during the IGY and continued after the conclusion of the 18-month period is a case in point.

The geophysicist cannot work without observational data and these require, in many cases, very expensive and difficult logistics. Rockets, satellites, ships for oceanographic expeditions, and high-altitude balloons for meteorological observations are but a few of the tools required.

Geophysics is a science whose investigations must be carried out on a world-wide basis. Thus the International Council of Scientific Unions (ICSU), with its 45-nation membership and 13 unions, which sponsored the IGY, continues to bear great responsibility for careful planning and execution of the research program. In October, 1958, the ICSU met in Washington to plan for post-IGY activities. It established a new Special Committee for Inter-Union Cooperation in Geophysics to carry on after the IGY, and to guarantee continuance of international collaboration in geophysics. In addition, the ICSU established a number of special committees in areas of geophysics which had to be continued on the same or an even larger scale than during the IGY.

Specialized planning bodies. The Special Committee on Oceanic Research (SCOR), the Special Committee on Antarctic Research (SCAR), the Committee on [illegible] Terrestrial [illegible] pace [illegible] vice [illegible] ttees [illegible] is of the world, particularly the larger ones, the academies of sciences have established national committees to correspond with these special committees.

A notable characteristic of both the special committees of the ICSU and of the national academies

is their breadth. Recognizing that geophysics is closely linked to other fields of science, these new committees are remarkably broad in their membership and in this way give promise of developing closer ties between the environmental sciences and other disciplines, such as physics, chemistry, and the life sciences.

Intra- and interdisciplinary aspects. These post-IGY plans also give promise of a greater intellectual challenge and usefulness to geophysics. Examples of enhanced interdisciplinary activity are the increasing number of studies involving the sun and the earth, the areas referred to as solar-terrestrial studies. After this interdisciplinary revolution inside geophysics a similar development may follow in which geophysics interacts with other sciences.

It is difficult to describe or to list all of the institutions interested in geophysics. In planning the IGY, an attempt was made to bring in, either directly or indirectly, all of the institutions and individuals who might help in the process. This was true also in the operational phase of the program and in the large number of post-IGY activities that have taken place.

Estimating future trends. It seems evident in viewing the future that activities in geophysics will continue to be greatly extended. Added to well-known departments and institutions will be new groups and individuals whose interests and talents have been diverted to this expanded and challenging field.

The IGY was built on a long tradition of international cooperation in geophysics By extending international cooperation, geophysics has excited the thinking of men in all fields and has done much to generate hope for peaceful cooperation in many other fields. [J.K.]

Geostrophic wind

A hypothetical wind based upon the assumption that a perfect balance exists between the horizontal components of the Coriolis force and the horizontal pressure gradient force per unit mass, with the implication that viscous forces and accelerations are negligible. Application of the geostrophic wind facilitates an approximation of the wind field from the pressure data over vast regions in which few wind observations are available.

Bases of the approximation. The geostrophic wind blows parallel to the isobars (lines of equal pressure) with lower pressure to the left of the direction of the wind in the Northern Hemisphere and to the right in the Southern Hemisphere. Its speed is given by

$$V_{geo} = \frac{1}{2\rho\Omega \sin\phi} \frac{\partial p}{\partial n}$$

where ρ is the density of air, Ω is the angular speed of rotation of the earth, p is the atmospheric pressure, ϕ is the latitude, and the n is a coordinate normal to the isobars and directed toward higher pressure. The approximation now known as the geostrophic wind was first derived empirically by C. H. D. Buys-Ballot in 1857 and has been known as Buys-Ballot's law.

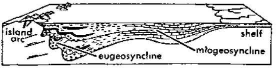
The geostrophic wind represents a good approximation to the actual wind at elevations greater than about 3000 ft, except in instances of strongly curved flow and in the vicinity of the Equator.

Thermal wind. This is a term denoting the net change in the geostrophic wind over some specific vertical distance. This change arises because the rate of change of pressure in the vertical is different in two air columns of different air density, so that the horizontal component of the pressure gradient force per unit mass varies in the vertical. The thermal wind is directed approximately parallel to the isotherms of air temperature with cold air to the left and warm air to the right in the Northern Hemisphere, and vice versa in the Southern Hemisphere. Thus, for example, the increasing predominance of westerly winds aloft may be viewed as a consequence of the warmth of tropical latitudes and the coldness of polar regions. *See* CORIOLIS ACCELERATION AND FORCE; GRADIENT WIND; WIND; WIND STRESS. [F.S.]

Geosyncline

A part of the surface of the earth that sank deeply through a long interval of time; ideally a great trough hundreds of miles long and tens of miles wide that received thousands of feet of sedimentary and volcanic rock through millions of years. A synclinorium is, in contrast, a great downfold that was produced by deformation later than the deposition of the contained rock. It has smaller flanking synclines and anticlines. *See* ANTICLINE; SYNCLINE.

Linear geosynclinal belts, or orthogeosynclines, are major tectonic divisions of the crust of the earth. They lie between more stable areas, or cratons, the higher continents, and the lower ocean basins. Orthogeosynclines generally have volcanic geosynclines toward oceans (called eugeosynclines or "true" geosynclines) and nonvolcanic miogeosynclines toward the continental cratons (*see* ORDOVICIAN). The eugeosynclines have been deformed in orogenies and intruded by igneous rocks that metamorphosed their sedimentary and volcanic rocks; they are analogous to modern volcanic archipelagoes such as the East and West Indies. The deep oceanic troughs associated with these archipelagoes are geosynclines that have not been filled, or leptogeosynclines



Diagrammatic section of Cordilleran geosyncline in southeastern Alaska and British Columbia at the close of the Permian. Volcanic deposits indicated in b' *(After A. J. Eardley, J. Geol., 55-319–342, 1947)*

Tectonic mobility of subsiding geosynclines is commonly complemented by the rise of welts, forming geographic island chains. Thus geosynclines commonly have terrigenous sediments eroded from such lands. But it is axiomatic that geosynclines are the sites of deposition of the greatest thicknesses of many kinds of surficial sedimentary and volcanic rocks, as such thickness leads to their classification as geosynclinal. It is doubtful that there is a geosynclinal cycle of deposition, for the sediments relate to factors beyond the site of deposition.

In addition to the linear orthogeosynclines, there are several kinds of related structural depressions, some doubtfully geosynclines. The thickest terrigenous sediments in the Appalachian region were laid in elliptical exogeosynclines within the cratonal margin, having been eroded from mountains built in the orthogeosynclinal belt to the east. Triassic rocks were laid in fault-bounded taphrogeosynclines along the Atlantic Coast from Nova Scotia to the Carolinas. Sediments, two miles or more in thickness, were laid in Pennsylvanian time in Colorado in zeugogeosynclines associated with intracratonal uplifts. The Gulf Coast geosyncline subsided nearly 10 miles during late Mesozoic and Cenozoic along the north coast of the Gulf of Mexico. Other subsiding intracratonic basins independent of rising lands, the autogeosynclines, have some aspects of geosynclines; Middle Paleozoic rock in southern Michigan filled such a basin. Rates of subsidence commonly are about 1000 ft in 1,000,000 years for tens of millions of years, but range to a maximum of about 3000 ft in 1,000,000 years.

Though some geosynclines have not been subjected to mountain making, orogenies are virtually restricted to geosynclinal areas, and severe orogenies with massive intrusions, to eugeosynclinal belts. *See* OROGENY. [M.K.]

Bibliography: M. Kay, *North American Geosynclines*, Geol. Soc. Am., Mem. 48, 1951.

Geraniales

An order of the plant subclass Dicotyledoneae, in which relationships are uncertain. In the Engler-Prantl system, the order includes 14 families having 654 genera and about 14,670 species. Usually the stamens are in two whorls and the ovules are pendulous. The members of the group have a wide distribution. There are many ornamentals such as geranium, crown-of-thorns, oxalis, nasturtium, croton, and poinsettia. Economically important members are flax, coca (a source of cocaine and an extract used in Coca-Cola), lignumvitae (hardest and heaviest of commercial woods), citrus with its many varieties, mahogany, the rubber tree (*Hevea brasiliensis*), tung-oil tree, and castor-oil plant. See separate articles on the economically important members listed in this article; *see also* DICOTYLEDONEAE; EMBRYOPHYTA; PLANT KINGDOM; TREE. [P.D.S.]

Germ

A primary source, especially one from which growth and development is to be expected. In biology the term denotes the substance of an egg or ovum or the earliest stage of any organism. It is used loosely to designate a microorganism such as a bacterium or a virus, without reference to its exact scientific classification, either directly, as in lay discussions of disease, or in compound terms such as "germ-warfare." *See* BIOLOGY; EMBRYOLOGY; MICROBIOLOGY. [H.P.T.]

Germ layers

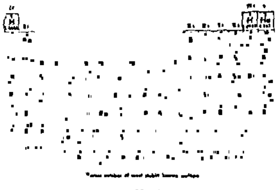
The primitive cell layers or first tissues, which appear early in the development of animals, and from which the embryo body and its auxiliary membranes, when present, are constructed. The individual cell units of the embryo are organized into supracellular units, the germ layers. These are more or less distinct anatomically, but do not necessarily have sharp boundaries of demarcation. Germ layers are almost universal among animal embryos and would appear to be an efficient method for establishing discontinuities of architectural importance without complete loss of continuity. Three kinds of germ layers are recognizable: (1) the ectoderm or outer skin, (2) the endoderm or inner skin, and (3) the mesoderm or middle skin. The layers have thus been named in accordance with their definitive positions in the spherical type of gastrula such as that of the sea urchin or amphibian. In gastrulae of reptiles, birds, mammals, and a few invertebrate types, where the spherical form has been modified into a blastoderm (a two-layered disk equivalent to a flattened sphere), the definitive positions of the layers remain essentially the same. The terms epiblast, mesoblast, and hypoblast are sometimes used as synonyms for ectoderm, mesoderm, and endoderm, respectively. The three primary germ layers are present as a basic structural plan in all Metazoa except the coelenterates and the Porifera, where a distinct mesodermal layer is absent.

Origin of the germ layers. The germ layer structure of embryos has been known for over a hundred years. In 1817 Heinrich Christian Pander described the trilaminar structure of the chick blastoderm, and a few years later Karl Ernst von Baer (1828–37) recognized that the layer concept held true for many types of embryos, both vertebrate and invertebrate. During the latter half of the nineteenth century, the following concept of the origin of the germ layers gradually developed. The blastula was considered to be a single-layered, hollow sphere, which became converted into a two-layered gastrula by a process of invagination or delamination of cells from one wall of the blastula. The outer layer thus became the ectoderm, the inner layer, the endoderm. A third layer, the mesoderm, then arose from part of either the inner or outer layer, depending upon the species of animal.

mester of pregnancy, some recommend therapeutic abortion. *See* ANIMAL VIRUS; ANTIBODY. [J.L.M.]

Germanium

A brittle silver-gray element (symbol Ge) with properties intermediate between those of silicon and tin. It is distributed widely in the earth's crust as the sulfide or associated with sulfide ores of other elements. The major use of germanium is in the manufacture of transistors, diodes, and rectifiers.



In 1864 John A. R. Newlands noted that there was a missing element between silicon and tin in the periodic table. In 1871 D. I. Mendeleev accurately predicted many of the properties of this element from a study of the properties of adjacent elements in the periodic table. Germanium was finally isolated in 1886 by C. A. Winkler and named after his native country of Germany. The first commercial production of the free element began in 1942.

Occurrence. Winkler first isolated germanium from the rare silver mineral argyrodite, $4Ag_2S \cdot GeS_2$, which was obtained from a mine in Freiberg, Germany. In the United States, the principal source is the zinc ores of the Tri-State district of Missouri, Kansas, and Oklahoma. The zinc sulfide mineral in these deposits contains 0.01–0.1% germanium. The recoverable germanium in this district may total more than 100 tons. The 1958 price was $170 per pound.

The zinc ore rock is treated by conventional ore-dressing methods to obtain a zinc sulfide concentrate. This concentrate is roasted to form crude zinc oxide which is sintered with salt and coal at high temperatures. The sintering operation vaporizes germanium, cadmium, and other impurities. The vapors are recovered in electrostatic precipitators. This recovered material is dissolved in sulfuric acid, and the germanium is concentrated in the residue, which is treated further to reduce the concentration of lead. The prepared residue is heated with excess strong hydrochloric acid to distill germanium tetrachloride. This raw germanium tetrachloride is further purified and then hydrolyzed in water to yield germanium dioxide.

Germanium is produced by hydrogen reduction of the dioxide at 675°C. Further purification is accomplished by using W. G. Pfann's zone-refining process.

In England and Russia, germanium is recovered from flue dusts obtained from the manufacture of producer gas. Some British coals contain about 0.003% of germanium, which remains in the coke produced from the coal. Combustion of the coke under limited supplies of air, as in the manufacture of producer gas, results in the germanium being driven off with the gas. When the hot gas is burned, metallic constituents are converted into oxides which are deposited in the cooler parts of the flue system. These flue dust deposits may contain as high as 2% germanium, and in England they constitute the richest sources of germanium, as well as of gallium. Production of germanium was started in 1950 using flue dust as the primary raw material.

Germanium occurs in certain ores in the Belgian Congo as the mineral renierite, a complex copper-iron-germanium-arsenic sulfide with an average content of 7% germanium. This is the most promising world source for germanium. Germanium concentrate is also prepared from copper concentrate in South-West Africa. Pure germanium oxide and germanium are produced from these germanium concentrates in Belgium.

Physical properties. Germanium crystallizes in the diamond structure with a lattice constant of 5.657 A. When pure, it is transparent to infrared radiation of wavelengths greater than 2 microns (μ). The index of refraction is 4.116 at 2 μ and 4.072 at 2.50 μ. The dielectric constant is 16. The vapor pressure is 10^{-2} mm Hg at 1420°C and 10^{-6} mm Hg at 950°C.

Germanium's electrical properties, first noticed in 1922, place it in the class of materials known as semiconductors. One characteristic of pure semiconductors is that their electrical resistivities decrease as the temperature increases. The resistivity depends upon the quantity and type of impurities present and the temperature. For pure material the resistivity is 60 ohm-cm at 25°C.

Chemical properties. At 25°C germanium is very stable; it is not greatly affected by water, 50% sodium hydroxide, concentrated hydrochloric acid,

Table 1. Properties of germanium

Atomic number	32
Atomic weight (natural mixture of isotopes)	72 60
Isotopes and abundance	
Mass 70	20 4%
Mass 72	27.4%
Mass 73	7.8%
Mass 74	36 6%
Mass 76	7 8%
Electron configuration	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^2$
Hardness, Mohs scale	6–6¼
Density, g/cm³	5 323
Melting point	936°C
Latent heat of fusion, cal/g	114 3
Specific heat at 25°C, cal/(g)(°C)	0 086
Thermal conductivity, cal/(cm²)(cm)(sec)(°C)	0 14

dilute hydrochloric acid, or sulfuric acid. At 100°C, sulfuric acid will slowly dissolve germanium. Nitric acid–hydrofluoric acid mixtures and molten alkalies, such as sodium hydroxide, quickly dissolve germanium.

Germanium compounds. Germanium in its compounds is either divalent or tetravalent. The divalent compounds are easily reduced or oxidized. The tetravalent compounds are more stable.

Halogen compounds. The germanium halides, germanium tetrachloride ($GeCl_4$), germanium bromide ($GeBr_4$), and germanium iodide (GeI_4), can be prepared by the direct reaction of the corresponding halogen with germanium. Germanium tetrafluoride, GeF_4, is prepared by the thermal decomposition at 600°C of barium fluogermanate, $BaGeF_6$, in quartz containers. $BaGeF_6$ can be obtained by mixing saturated barium chloride solution with a solution of germanium dioxide in 48% hydrogen fluoride.

Every commercial method for the extraction of germanium from its ores involves the preparation of $GeCl_4$, which boils at 83.1°C and freezes at −49.5°C. A common impurity extracted at the same time is arsenic. The arsenic can be removed by distillation of a mixture of germanium and arsenic chlorides in the presence of hydrochloric acid and chlorine. $GeCl_4$ is immiscible with hydrochloric acid, which floats on the surface. Arsenic trichloride dissolves in the aqueous layer and is oxidized to arsenic acid, which is not volatile.

$GeCl_4$ enters into Grignard, Wurtz-Fittig, and similar reactions very readily. Like silicon, tin, and lead, germanium forms a large number of organometallic compounds.

Oxygen compounds. Two oxides of germanium are known, GeO and GeO_2. Three crystalline forms of GeO_2 have been observed.

The hexagonal form of GeO_2 is obtained by the hydrolysis of $GeCl_4$. This form is soluble in water, 0.45 g/100 g water at 25°C. It reacts with hydrochloric acid to yield $GeCl_4$ and with hydrofluoric acid to yield fluogermanic acid, H_2GeF_6.

Ignition of the hexagonal form of GeO_2 at 380°C alters part of the oxide to the tetragonal form, which is insoluble in water and does not react with hydrochloric or hydrofluoric acid.

An amorphous, transparent, glassy form of GeO_2 is formed by rapid cooling of melted GeO_2. This form does not have a definite melting point and is soluble in water. It devitrifies into an opaque white glass.

Hydrogen compounds. The properties of some germanium-hydrogen compounds are summarized in Table 2.

Table 2. Germanium hydrides

Name	Formula	Melting point, °C	Boiling point, °C
Monogermane	GeH_4	−165	−90
Digermane	Ge_2H_6	−109	29
Trigermane	Ge_3H_8	−105.6	110.5

The hydrides of germanium can be produced by the reaction of hydrochloric acid with magnesium germanide. This method yields a mixture of the germanes. C. A. Kraus obtained higher yields by the reaction of magnesium germanide with ammonium bromide in liquid ammonia. High purity (99.7%) monogermane can be obtained by the reaction of the germanes with sodium in liquid ammonia followed by the recovery of monogermane by treatment with ammonium bromide. All the germanes undergo thermal and photochemical decomposition.

Alloys. At 55 weight per cent germanium, the binary system germanium-aluminum forms a eutectic, which melts at 423°C. At 12 weight per cent germanium, the germanium-gold system forms a eutectic, which melts at 356°C. Silicon and germanium form a continuous series of solid solutions.

The relative solubilities of small amounts of boron, aluminum, gallium, phosphorus, indium, arsenic, antimony and several other elements in solid and liquid germanium are quite important in the preparation of germanium for semiconductor applications. *See* SEMICONDUCTOR.

Uses. During World War II, germanium was intensively investigated for its use in the rectification of microwaves for radar applications. Several types of diodes were developed. The production level of about 6,000,000 in 1951 increased to 41,209,000 in 1957.

A major development in electronics occurred in 1948 with the invention of the transistor. This solid-state amplifying device, which can be made of germanium, has had a profound influence on all electronic applications and has already captured the hearing-aid and portable radio markets. About 1,300,000 transistors were produced in 1954 and this increased to 27,706,000 in 1957.

Germanium lenses and filters are being used in instruments which operate in the infrared region of the spectrum. Glasses prepared with germanium dioxide have a higher refractivity and dispersion than do comparable silicate glasses. These glasses may be used in wide-angle camera lenses and in microscopes.

The magnesium germanate phosphor is used in some fluorescent lamps. Gold-germanium alloys which expand on freezing have been suggested for use in jewelry solder and for dental work. Germanium-silver film resistors have been developed with temperature coefficients less than 0.0001 per 1C°.

The pharmacological uses of germanium compounds have been investigated, but no important applications have as yet been developed.

Analytical chemistry. Samples are usually taken into solution by a sodium carbonate fusion followed by a cold water extraction. Germanium is separated from other elements by precipitation with hydrogen sulfide in a 6 *N* sulfuric acid solution; the germanium sulfide is dissolved in dilute ammonium h droxide, transferred to a distillation flask, and tilled from a 1:1 HCl solution under 1 atmos of chlorine.

Identification. Germanium may be detected in solution by using hydrogen sulfide, potassium ferrocyanide, hydrogen selenide in formaldehyde solution, phenylfluorone test paper, and several other organic reagents.

Determination. Germanium is commonly determined gravimetrically by precipitation with tannin followed by ignition to GeO_2. Germanium(IV) in aqueous solution reacts with mannitol to form a strong monoprotic acid complex which can be titrated with sodium hydroxide. In the presence of strong electrolytes, the germanium-mannitol complex liberates iodine quantitatively from a potassium iodide-potassium iodate solution. The liberated iodine can be titrated with sodium thiosulfate.

Microgram quantities of germanium are determined spectrophotometrically with phenylfluorone as the color-developing reagent. Germanium reacts with ammonium molybdate to form a soluble heteropoly germanomolybdate complex which has been used to determine germanium colorimetrically. The complex can be reduced with ferrous sulfate to yield molybdenum blue, which permits a more sensitive measurement of the germanium content. *See* RECTIFIER; [illegible]

Bibliogr. [illegible]
Handbook, [illegible]
their orga[illegible]
1951; O. [illegible]
compounds [illegible]

Germanium diode

A small germanium rectifier of either the point-contact, bonded contact, or small junction type. It is distinguished from a germanium rectifier by size only, the latter term being used for units of relatively large power-handling capacity. The point-contact and bonded or welded types have found wide application in television, computers, and microwaves. In general germanium diodes are limited to peak inverse voltages below 300 volts and to operating temperatures below 100°C. When properly used they show remarkably long life and stability. A steady improvement in characteristics has accompanied a trend toward microjunction techniques. *See* JUNCTION DIODE; POINT-CONTACT DIODE. [L.P.HU.]

Germfree vertebrate

An animal that is free from all other demonstrable forms of life. Within the limitations of present germfree techniques, rabbits, goats, guinea pigs, rats, mice, chickens, turkeys, and platyfish can be obtained free from bacteria, yeasts, fungi, protozoa, endo- and ectoparasites, and many of the viruses common to the species. The question of freedom from nonsymptomatic viruses, proviruses, and cytopathogenic viruses remains to be explored. *See* VIRUS.

It has been shown that some vertebrates can live normal spans of life and even reproduce under germfree conditions. Mice have been reared germfree through 12 generations and have survived for periods up to 900 days. Rats have been reared germfree through 14 generations and have survived as long as 900 days. Chickens have been reared germfree through 2 generations in limited efforts in this direction. These are the only species reared into successive generations under germfree conditions.

Germfree life studies are based upon an extension of the pure culture concept, hence it is now plausible to speak of pure cultures of, for example, mice or rats. On this basis the study and use of germfree animals takes on a new dimension. It also brings with it difficulties in terminology since it is often necessary to infect germfree animals with a single pure culture of microbes in which instance the term monocontaminant has been used. *See* CULTURE, PURE.

Apparatus. Germfree life studies are dependent upon apparatus in which a sterile environment can be maintained and animals can be reared and subjected to experimentation without introducing contamination. Two such systems of isolators are the Reyniers' systems 1 and 2 which have been operated and tested for over 25 years. These systems permit working with either large or small animals under a variety of experimental conditions.

System 1 is designed primarily for large animals such as goats, dogs, monkeys, and pigs. It is operated with steam and germicides and permits entry of an attendant clothed in a plastic garment.

System 2 consists of a series of isolators which can be connected together and operated through attached rubber gloves. Sterilization is effected by autoclaving and, in certain instances, with germicides. The basic isolator is conventionally a cylinder equipped with a pair of gloves, air filters, a viewing window, external lighting, and a supply lock.

The basic isolator may be modified in shape and size into three types: (1) rearing isolator, for rearing, breeding, and holding animals, (2) surgical isolator, for performing cesarotomies, and (3) examining isolator, usually equipped with two or more pairs of gloves, for experimental procedures involving special apparatus such as homogenizers, centrifuges, and filtering devices. The isolators may be joined together to form an experimental system in which many individuals may work together in the performance of a variety of procedures. Each isolator of the system, when connected by a supply lock, may be sterilized independently, and may be either taken out of the system or joined to it without contamination.

Food, bedding, water, and instruments may be introduced into an isolator, debris may be removed from it, or animals may be moved from one isolator to another by way of the supply lock. The supply of air is sterilized by a series of steps and introduced into the isolator through fiberglass in attached filters which are sterilized in position and dried by vacuum. All operations have been reduced to routine and can be performed by trained labora-

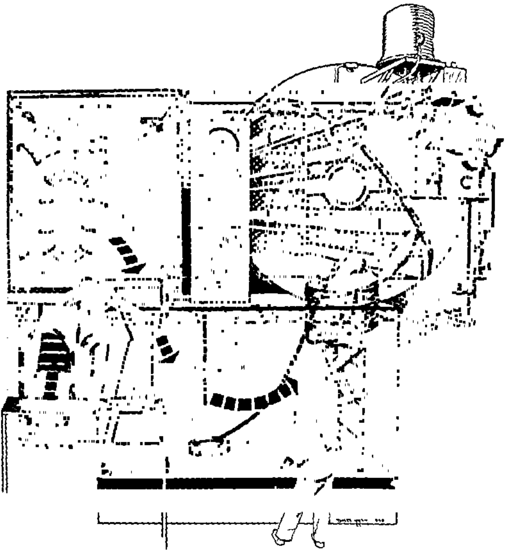


Fig. 1. Reyniers' system 1. An isolator for large-animal experimentation.

tory technicians. Single isolators have been used continuously, without changing the animals, for periods of over 2 years.

The isolators have been redesigned and are produced commercially with transparent vinyl plastic walls. The transparent isolator with its flexible plastic envelope makes it impossible to obtain a negative pressure within the chamber. By this means, a contamination hazard, common to rigid chambers, is eliminated.

Obtaining germfree vertebrates. Viviparous animals are obtained germfree by cesarotomies performed at term in a surgical isolator connected to a rearing isolator. Animal species dependent upon the mother at birth must be hand-fed a sterilized, or autoclaved, liquid diet until they are weaned. At maturity, they may be bred and will rear their young. Some mammals, like the guinea pig, can be reared by self-feeding on semisolid diets; others, like platyfish, must have the food particles in the supporting medium.

Oviparous animals, like chickens and fish, are obtained by sterilizing the shell or chorionic membrane of the fertile eggs with germicides and passing them through a germicidal trap into the isolator where they are incubated to hatching. These animals will feed themselves.

Testing for contamination. Tests for contamination are performed either on the living animal or on one of the group which is sacrificed for that purpose. Microbiological tests of intestinal contents, droppings, food, water, bedding, blood, and biopsy material can be performed without sacrificing the animal. These tests can be supplemented by serological examination. Special tests for protozoa and other animal parasites, and for viruses can be included in the testing routines. *See* SEROLOGY.

Characteristics of germfree vertebrates. It is difficult to generalize about germfree life since there are species differences and the information obtained has been largely incidental to the technical problems involved. There is a real difference be-

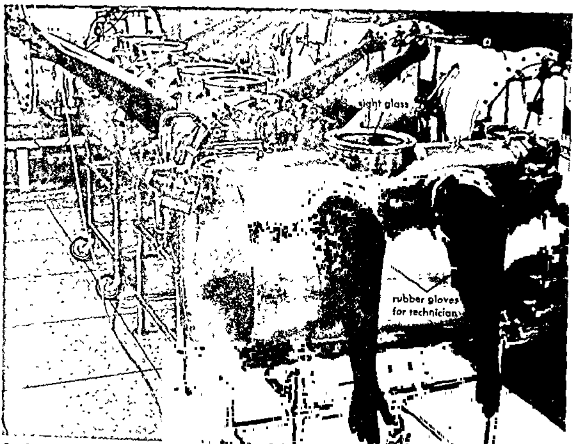


Fig. 2. Reyniers' system 2. Isolators for small-animal experimentation.

tween germfree and conventional life. The germfree animal has never had experience with living contaminants, whereas the conventional animal has been contaminated with a variety of microbes from the time of its birth. Nevertheless, the same systems are present in each, and the differences may be that certain systems are not activated in the germfree animal. How far this may be reflected in associated systems and in the general economy of the animal remains to be determined.

Higher vertebrates living in the absence of mi-

ment. These animals grow, mature, breed, rear their young, and pass into successive generations, thus meeting the biological demands for the species. So far as is known, these animals live as long, react to nutritional deficiencies, and show red cell and hemoglobin values close to normal for the species. They mature at about the same rate and produce a comparable number of litters and young per litter.

Differences are noted by the absence of fecal odors and post mortem changes in the germfree animal as compared to the conventional. There is no decay in the accepted sense of the word in germfree animals. Organs commonly in contact with bacteria, such as the gastrointestinal tract, exhibit differences. For instance, the gastrointestinal tract in germfree animals weighs less, and there is less lymphatic and connective tissue, a reduced blood supply, and in animals with a cecum, this organ is usually enlarged. The liver, however, appears to be smaller in the germfree animal. Systems and organs not in immediate association with microbes in the conventional animal have a similar appearance in the germfree vertebrate.

The serum proteins are low in gamma globulin in the germfree animal, and in the chicken it appears that there are fewer plasma cells and fewer secondary nodules (see GLOBULIN). Heterohemagglutinins are present in germfree animals, and properdin levels are demonstrable (see PROPERDIN). The germfree animal responds specifically to antigenic stimulation when it is introduced parenterally or orally. Resistance to infection depends on the species of microorganisms. For example, germfree mice of the C_3H strain are affected adversely by exposure to a conventional environment, but rats are not. Monocontaminants such as *Shigella dysenteriae* can be established in germfree mice and rats at high levels without apparently affecting the animals, whereas normally saprophytic microorganisms such as *Shigella lutea* and *Bacillus subtilis* may be pathogenic to germfree chickens. See SHIGELLA.

The diets used to rear germfree animals are necessarily sterilized, preferably by autoclaving. It is therefore necessary to make up losses in vitamins

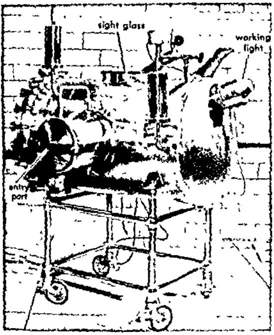


Fig. 3. Reyniers' basic isolator.

and other components destroyed by heating. The diets developed to date seem adequate to rear germfree or conventional animals. One of the problems with respect to diets occurs during the suckling period of those animals dependent upon the mother when these diets must be hand-fed as is the case with cesarean-born rats and mice. In general, semisynthetic diets are better suited to germfree rats and mice than are the cruder diets containing cellulose and natural foodstuffs. Germfree guinea pigs and chickens respond more favorably to diets containing natural grain products.

Uses of germfree apparatus. In addition to the aforementioned, germfree apparatus is useful in maintaining animals with known single or multiple contaminants. Since the apparatus is suited to environmental control, it can be used for ecological and physiological studies. It may also be used in chemical experiments wherein the need for controlling gas, temperature, and humidity is important. Such apparatus has been used in such operations as machining and welding where surface oxidation is undesirable, but where a great deal of manual activity is necessary By substituting mechanical arms for rubber gloves, high pressures can be employed, and it is possible to work with radioactive materials. The apparatus is used extensively to contain dangerous pathogens, such as in the study of air-borne infections induced by clouds of pathogens, and in the transmission of disease by insects.

Experimental biology and medicine. The study of germfree life constitutes a useful and important field in itself. Germfree animals may also be used as pure cultures or pure animals to study microbial associations. The following is a short list of some of the problems in which germfree life has been successfully employed and others in which it may be used.

1. The study of diseases of unknown etiology, for example, cancer, particularly Hodgkin's disease and leukemia.

2. As an experimental animal for growing the virus of hepatitis or *Shigella dysenteriae*, one of the bacteria which cause bacillary dysentery.

3 Nutrition studies, for example, synthesis of biologically labile methyl groups, the growth effect of antibiotics, biosynthesis of folic acid and citrovorum factor, vitamin interrelationships, intestinal synthesis. Intestinal synthesis implies the formation of vitamins and other accessory factors

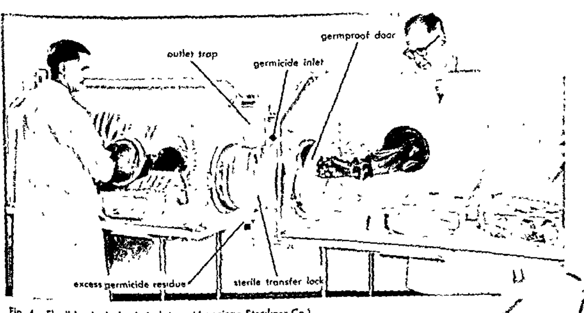


Fig. 4. Flexible vinyl plastic isolators. (American Sterilizer Co.)

by microorganisms in the intestinal tract of animals and utilization of these compounds by the animal.

4. The *etiology of dental caries.*
5. Experimental liver necrosis.
6. Relationship of microbes to the syndromes of irreversible hemorrhagic shock and in total body irradiation.
7. Ameba-bacterial relationships in experimental amebiasis.

See MICROBIOLOGICAL METHODS; RADIATION INJURY (BIOLOGY). [J.A.R.]

Gerontology

The study of aging of any material, usually restricted to those changes seen in biological systems, particularly in man

Gerontology includes at least three major, closely related subdivisions which deal with the cells and tissues of a given individual, the entire aging process of the whole individual, and the sociologic, economic, cultural, political, and other group problems of the elderly as a population segment.

Senescence. The first subdivision, the study of the aging processes of the cells and tissues of an individual, is known as the biology of senescence. The effects of advancing age may be seen as certain changes in organs, tissues, metabolic systems, and in the ultimate biologic unit of the cell itself. Many tissues, such as the skin and liver, show a great ability to regenerate, to withstand the repeated insults encountered during the passage of time, and show degenerative changes, under normal conditions, much less than certain other body structures. The joints, bones, connective tissue, and body fluids are principally composed of nonliving *structures*; that is, they are relatively acellular. These structures show a decided tendency to de*generate with* advancing age, although cellular structures and organs are certainly not spared in later life Understanding of the physical, chemical, and biological properties and interrelationships of the body components sheds much light on *the entire field* of gerontology and also offers specific content for preventive or remedial approaches to certain problems of aging *See* SENESCENCE.

Geriatrics. The study of the individual as a whole is called geriatrics. More specifically, in terms of physical and mental health, the recently developed field of geriatric medicine considers appropriate problems. The fact that the population in the United States has shown a progressive increase in numbers of elderly persons has caused attention to fix on many medical problems *not so* widely encountered in younger groups. Although much stress is given to those diseases and disorders preponderant in elderly persons, such as certain forms of joint, heart, kidney, gastrointestinal, and *metabolic disturbances*, the psychic or emotional components are receiving increasing attention. More important, perhaps, is the fact that much investigation has pointed to the desirability of utilizing specific methods in preventive medicine which can be most beneficially applied *before the onset of old age*. In other words, it is increasingly recognized that the best care does not wait until the irreversible changes of aging become manifest, but rather attempts to prevent or delay their onset.

Aging populations. The third aspect of gerontology, the consideration of large groups or populations of elderly people, is inseparable from the other aspects. This over-all view may be broken up into portions by the observer so that he may consider, for example, only the elderly population of a neighborhood or a town, or of the personnel in a business. This approach provides, of course, enlarged details of nationwide or world-wide problems that have been created by the fact that there is an increasing percentage of older people in most parts of the world.

Few segments of life are free of the considerations included in the study of gerontology, whether the individual is concerned with farm production, government administration and finance, or military planning, or concerned with individual, family, or community considerations.

When arrangements are made for periodic physical checkups, pensions, savings or for health, accident, or life insurance, it is an attempt to hold off some of the deleterious effects of aging so that the benefits can be more fully enjoyed. [E.C.ST.]

Gestation period

That period from fertilization of an egg to birth, in the mammal. In man, gestation lasts from 270 to 295 days, or roughly 9 calendar months. Accordingly, for clinical purposes, this period is divided into three equal trimesters. The fertilized ovum becomes implanted in the highly vascularized uterine mucosa within a week of fertilization and the early embryo and placental membranes begin to form

In other mammals the gestation period varies considerably, as shown in the table, and is dependent upon such factors as maternal care, metabolic rates, environmental conditions, and types of placentation. *See* FERTILIZATION; REPRODUCTIVE SYSTEM; REPRODUCTIVE SYSTEM DISORDERS.

Gestation periods of common mammals

Animal	Gestation, days	Animal	Gestation, days
Armadillo	150	Guinea pig	68–71
Bear		Horse	330–380
Black	210	Kangaroo	40–45*
Polar	240	Lion	106
Cat	60	Man	270–295
Chimpanzee	250	Mole	30
Cow	282	Mouse	20–21
Dog	58–65	Opossum	13*
Donkey	365–380	Rabbit	30–43
Elephant		Whale	331–365
African	641	Wolf	63
Indian	607–641	Zebra	300–345
Giraffe	450		

* Marsupial, with extended stay in maternal pouch

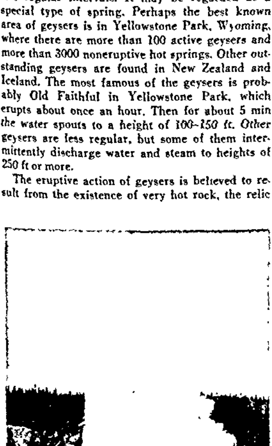
By the 47th day, the human embryo has developed all the characteristic organs and tissues of man. From this time on, the term fetus is used Sometime during the second trimester of pregnancy spontaneous movement will be felt. Loss of the embryo or young fetus prior to the stage of viability is known as abortion; later loss is called premature delivery until the fetus weighs more than 2500 g, the lower limit of normal weight of the newborn.

Growth during the latter half of pregnancy is characterized by an increase in size and weight and few basic alterations appear in the fetus. [E.C.ST.]

Geyser

A natural spring or fountain which discharges a column of water or steam into the air at more or less regular intervals. It may be regarded as a special type of spring. Perhaps the best known area of geysers is in Yellowstone Park, Wyoming, where there are more than 100 active geysers and more than 3000 noneruptive hot springs. Other outstanding geysers are found in New Zealand and Iceland. The most famous of the geysers is probably Old Faithful in Yellowstone Park, which erupts about once an hour. Then for about 5 min the water spouts to a height of 100–150 ft. Other geysers are less regular, but some of them intermittently discharge water and steam to heights of 250 ft or more.

The eruptive action of geysers is believed to result from the existence of very hot rock, the relic of a body of magma, not far below the surface. The neck of the geyser is usually an irregularly shaped tube partly filled with water which has seeped in from the surrounding rock. Far down the pipe the water is at a temperature much above the boiling point at the surface, owing to the pressure of the column of water above it. Its temperature is constantly increasing, owing to the volcanic heat source below. Eventually the superheated water changes into steam, lifting the column of water out of the hole. The water may overflow gently at first, but as the column of water becomes lighter a large quantity of the hot water may flash into steam, suddenly blowing the rest of the column out of the hole in a violent eruption. [A.N.S.]



Old Faithful. (National Park Service, U.S. Dept. of Interior)

Ghost image, optical

An undesired image appearing at the image plane of an optical system. Each surface of an optical system divides the incoming light into two parts: (1) the reflected light, which returns into the first medium, and (2) the refracted light. The reflected light is again divided into two parts when it in turn strikes another dividing surface. The light thus reflected twice forms an image which may be near the plane of the primary image. This may be a false image of the object or an out-of-focus image of a bright source of light in the field of the optical system. Thus a large number of undesired or ghost images may appear. *See* IMAGE, OPTICAL; REFLECTION (ELECTROMAGNETIC RADIATION); REFRACTION OF WAVES.

If the ghost images are far out of focus, they only diminish the contrast in the primary image, a condition known as flare. But if the ghost images are near the focal plane, they are very disturbing. This effect is especially noticeable if there is a bright light source in the field of the instrument, since the ghost image of the light source may have an even greater brightness than the image of the desired object.

This annoying image defect is hardly precalculable because in a system with N surfaces there are 2^N possible double reflections and 4^N possible quadruple reflections. It has, however, been successfully controlled by antireflection coatings.

Coatings. Lens coatings are films of the proper thickness and refractive index applied to the air-glass surfaces of a lens to reduce reflection. For perpendicular incidence on a plane surface bounding a medium of refractive index n, a surface coating with an effective index of $\sqrt{n}$ and a thickness of a quarter of a wavelength causes all light waves of that wavelength returning into the first medium to interfere with each other destructively so that there is no loss by reflection. The coating of lenses with layers of fluorite and other materials has nearly eliminated ghost images from modern optical systems. Different coatings can be combined effectively as color filters for either transmission or reflection. They also can be used as interference filters in interferometry or spectroscopy. *See* INTERFERENCE FILTER, OPTICAL.

Flare. There are several causes of flare. One is reflection at an even number of lens surfaces which, especially in the case of an illuminated object, may give rise to an out-of-focus image of the light source that will produce background illumination. This type of flare, like ghost images, can be practically eliminated by suitably coating the air surfaces of the lens elements. Flare is also caused by reflection at the mounting of an optical instrument, but this can be eliminated by blackening and suitable baffling. A different type of flare arises from the fact that not all the emergent light is collected within the small nucleus of the image. This flare is due to residual aberrations that exist even in a well-corrected lens

Flare is especially detrimental if the object has small contrast differences, since it reduces these differences in the image and may bring them below the threshold of recognition On the other hand, by lightening the shadows of the image of a high-contrast scene, some background illumination, as produced by flare, may enable a photograph to be made with a shorter exposure than would be possible in the absence of it and may also bring the contrast range of the scene within the range of reproduction of the photographic process. [M.H.]

Bibliography: K. Blodgett, Extinction of reflection from glass, *Phys Rev.*, 55(4):391–401, 1939; J. Strong, *Concepts of Classical Optics*, 1958.

Giant star

A member of one of the families into which stars are divided in the Hertzsprung-Russell diagram; the term giant refers to true brightness rather than to size (*see* STAR). Although giant stars are frequent among the brighter, naked-eye stars, such as Arcturus and Capella, they are comparatively rare in space. They have luminosities approximately 40–200 times that of the Sun, and occur most commonly in a range of temperature from 7000 to 2500°K; a few hotter giants are known. The common red giant stars have radii about 10 times that of the Sun, masses from 1.5 to 4 times that of the Sun, and mean densities about 10^{-3} g/cm^3. They represent a comparatively late stage in stellar evolution. Some cooler red giants vary in brightness periodically. [J.L.GR.]

Gibberellin

One of a complex mixture of plant-growth-promoting substances produced by the fungus *Gibberella fujikuroi*. This fungus causes a disease in rice plants known in Japan as the bakanae, or foolish seedling, disease. One of the characteristic symptoms of diseased rice plants is that they grow taller than healthy plants. These observations led to the discovery in 1926 by E. Kurosawa, a Japanese plant pathologist working in Formosa, that the fungus grown in liquid culture secretes materials into the medium which accelerate the growth of rice plants. *See* ASCOMYCETES.

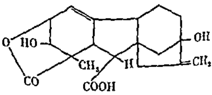
In 1938 two crystalline gibberellins, A and B, were isolated from the fermentation broth by a group of Japanese scientists, T. Yabuta, Y. Sumiki, and T. Hayashi. They also determined many of the chemical and biological properties of gibberellins and found that many other kinds of plants were stimulated to grow taller when treated with gibberellins. For the isolation work, the fungus was grown in a medium containing glycerol, ammonium chloride, and potassium dihydrogen phosphate. The active gibberellins were removed from the fermentation broth by adsorption on activated charcoal, and then eluted from the charcoal with methyl alcohol. The crystalline gibberellins were obtained when the methyl alcohol solution was concentrated.

Gibberellin A was converted into gibberellin B by mild heating in dilute acid. A third substance, gibberellin C (a monohydrate of A), was produced by boiling A in dilute mineral acid. Degradation of A or B yielded *gibberene*, a fluorene derivative.

Very little work on gibberellins was done outside Japan until 1954 when the British workers J. Curtis and B. E. Cross reported on the isolation of a new product, gibberellic acid ($C_{19}H_{22}O_6$). Yields of up to 180 milligrams per liter were obtained by growing the fungus in Raulin-Thom medium. Gibberellic acid has two less hydrogen atoms than gibberellin A. At about the same time that the British workers found gibberellic acid, a group of scientists in the United States Department of Agriculture headed by F. H. Stodola, published a method of producing gibberellin A and gibberellic acid. They grew the fungus in a medium containing glucose, ammonium chloride, magnesium sulfate, and potassium dihydrogen phosphate.

In 1955 the Japanese workers led by N. Takahashi reexamined their gibberellin A using more modern methods of separation and found that it contained three gibberellins which they named A_1, A_2, and A_3. Gibberellin A_3 was shown to be identical with gibberellic acid.

In 1956, Cross and his coworkers proposed a tentative structure for gibberellic acid (gibberellin A_3).



Gibberellic acid

The structure of gibberellin A_1 is the same as shown for gibberellic acid except that the double bond in the first ring is reduced. Recently another gibberellin, called A_4, was isolated by the Japanese scientists, so that four biologically active gibberellins, A_1, A_2, A_3, and A_4, are now recognized as products of the fungus *G. fujikuroi*. Additional ones will undoubtedly be discovered in the future. Gibberellins B and C and many esters and salts of the gibberellins also promote stem elongation,

but these can be produced from the original gibberellins by various chemical reactions.

Occurrence in higher plants. Gibberellinlike factors were first isolated from higher plants by C. A. West and B. O. Phinney in 1956. In 1958 they isolated, from bean seeds, two crystalline gibberellinlike factors which were very similar to gibberellins A_1 and A_3 in many respects but differed slightly in certain light-absorption characteristics.

J. MacMillan and P. J. Suter in 1958 were successful in isolating pure gibberellin A_1 from immature seeds of the runner bean. This finding established the natural occurrence of one gibberellin in higher plants, thus permitting the definition of gibberellins as plant-growth hormones, so that they can be added to the list of plant-growth hormones such as 3-indoleacetic acid and other auxins which occur naturally in higher plants.

Effects of gibberellins on plants. When solutions of gibberellins are sprayed on the foliage or applied to the roots of many kinds of plants, the most pronounced response is stem elongation which is noticed a few days after treatment. This increase in stem length results from both an increase in cell elongation and an increase in cell division. The size of the leaves of certain plants is also increased, and the shape of the leaves may be altered. Dwarf peas and certain dwarf corn varieties appear as normal or tall varieties after treatment. An increase in dry weight of the plants occurs in many stimulated plants.

Gibberellins can replace the cold requirement for flowering of certain long-day plants and can induce flower formation in some long-day plants grown under short-day conditions. Flowering of some plants is delayed by treatment, whereas it is speeded up in other species of plants. The size of the flowers is increased in some plants such as geraniums when the buds are treated.

Other responses caused by gibberellins are increasing fruit setting of tomato; breaking dormancy of freshly dug Irish potatoes; breaking dormancy of peach trees and peach seeds; breaking dormancy of certain seeds such as crab grass, wild oats, and lettuce; hastening germination and emergence of seedlings when seeds are treated; making pasture and lawn grasses grow in cold weather; increasing size of grapes and other fruits; increasing yields of crops such as celery and cotton Some of these applications have been adopted as sound agricultural practice, and other uses of gibberellins in agriculture are being developed.

Gibberellins are now produced commercially by drug companies by culturing the fungus in large fermentation tanks in a manner similar to that used for the production of antibiotics such as penicillin. The purified gibberellins are sold on the market and can be purchased in various formulations for garden and agricultural use. See PHOTOPERIODISM IN PLANTS; PLANT GROWTH; PLANT HORMONES.

[R.A.GR.]

Bibliography: Frank H. Stodola, *Source Book on Gibberellin 1828–1957*, USDA Agricultural Research Service; Bruce B. Stowe and Tashio Yamaki, The history and physiological action of the gibberellins, *Ann. Rev. Plant Physiol.* 8:181–216, 1957.

Gibbs function

The thermodynamic function g for free energy

$$g = h - Ts = u + \frac{pv}{J} - Ts$$

where g is the Gibbs function, h is enthalpy, T is thermodynamic temperature, u is internal energy, p is pressure, v is volume, s is entropy, and J is the heat equivalent of work. The function is particularly important in processes involving chemical change. For reversible isothermal steady-flow processes or for reversible constant-pressure isothermal nonflow processes, a change in free energy equals net work. See ENTROPY; THERMODYNAMIC PRINCIPLES.

[C.A.H.]

Gila monster

Heloderma suspectum, which is also called the beaded lizard, the only poisonous lizard found in the United States. A closely related species, the Mexican beaded lizard, *H. horridum*, occurs in western Mexico and northern Central America.

The gila monster, *Heloderma suspectum*; length to 24 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

These animals have grooved teeth on the lower jaw which are connected with venom glands, and they must bite with the back teeth and turn over to administer a severe dose of poison. They are dangerous to man, their venom being similar in action to that of poisonous snakes.

The Gila monster is easily recognized by its beadlike scales, its short, thick tail, and brilliant markings; the base color varies from pink to orange or yellow, marked with irregular blotches of black. It is found in the Southwestern desert from Utah and Nevada into northern Mexico. See SQUAMATA.

[J.D B.]

Gilbert

A centimeter-gram-second (cgs) electromagnetic unit of magnetomotive force. A gilbert is the magnetomotive force of a closed loop of one turn in which there is a current of $1/4\pi$ abamp ($10/4\pi$ amp). A gilbert is $10/4\pi$ amp-turns. See AMPERE-TURN; ELECTRICAL UNITS; MAGNETOMOTIVE FORCE.

[K.V.M.]

Gill (anatomy)

The respiratory organ of water-breathing animals. Gills are composed of various arrangements of thin, highly vascularized membranes over which water passes, thus permitting exchange of oxygen and other solutes. Gills develop from the branchial and pharyngeal pouches so that a series of slits are formed with intervening bars or septa. Located on either side of the neck, these septa then develop the thin membranes in a manner and number characteristic of the various orders and species. In most fishes, the number of gills has decreased from that of primitive vertebrates, and a lateral, fluctuating flap or operculum is present. In addition, many fishes supplant gill respiration by the action of an air bladder in the body which accumulates oxygen from the blood for future needs. *See* SWIM BLADDER

External gills are present in larval amphibians, but these are lost during metamorphosis.

All higher vertebrates, including man, show some degree of recapitulation of gill formation during embryonic development, and certain structures are derived from these primordial areas. For example, the inner ear, the larynx, the tonsils, and other structures are derived from the primitive gill anlage. *See* RESPIRATORY SYSTEM. [E.C.ST.]

Gilsonite

An asphaltite that occurs in the Uinta Basin of northeastern Utah. Most deposits lie in Uintah and Duchesne Counties, Utah. The nearly parallel veins have a northeast-southwest trend, are characterized by nearly vertical, sheer sandstone walls, and may be traced for distances up to 40 miles by surface outcrops. Exposures may range from several inches to 22 ft in width. The gilsonite is essentially free of mineral matter, but veins do occasionally contain floated blocks of wall rock. *See* ASPHALT AND ASPHALTITE.

Gilsonite is used in the manufacture of paints, battery boxes, asphalt floor tiles, brake linings, printing inks, and electrical insulation. It is also used for waterproofing and insulating high-temperature piping.

Gilsonite is black and has high luster, coarsely conchoidal fracture, and a red-brown streak. The material is friable and on mining or crushing breaks to a fine, penetrating dust. It is sold as selects, coming from the interior of a vein, or as seconds, obtained near the contact of the vein with the enclosing rock. The seconds are not as lustrous as the selects, have a higher fusing point and lower solubility in naphtha, and differ in specific gravity as shown in the following list.

	Selects	*Seconds*
Specific gravity at 77°F	1.03–1.05	1.05–1.09
Fusing point	230–240°F	240–280°F
Solubility in naphtha	50–60%	20–50%

The larger deposits of gilsonite have been mined since shortly after their discovery in 1882. Among the better known are the Cowboy, Bonanza, Rainbow, and Black Dragon veins. Some veins have been mined to nearly 1000 ft in depth.

The ultimate analysis of gilsonite is variable: carbon, 85–86%; hydrogen, 8.5–10.0%, sulfur, 0.3–0.5%; nitrogen, 2.0–2.8%; and oxygen, 0–2%. Gilsonite contains a small percentage of hydrocarbons (4–8%) that can be isolated by chromatography (*see* CHROMATOGRAPHY). Some of these compounds are naphthenic (60–70%), and others are aromatic (30–40%). Other compounds that have been isolated in very small quantities are unsaturated and have the following approximate empirical formulas: $C_{18}H_{25}O_3N_2$, $C_{10}H_{13}O$, and $C_{63}H_{95}N$. Gilsonite from the Bonanza vein has been found to contain 0.03% porphyrins; the nickel complex of deoxophylloerythroetioporphyrin has been isolated and identified.

Neither the origin of gilsonite nor its peculiar occurrence is completely understood. The material probably originated in sediments of upper Green River age (Eocene) as a lacustrine sapropel derived from both animal and vegetable sources (*see* SAPROPEL). The sapropel and its subsequent diagenesis to gilsonite were undoubtedly dependent upon the nature of the contributing organisms and the controlling calcareous and dolomitic environment of deposition (*see* DIAGENESIS). Thermal effects do not appear to have been important in the conversion of the sapropel to gilsonite. Gilsonite has been batch-distilled to temperatures of 660°–770°F yielding 50–55% distillate, 15–20% gas, and 30% coke. Analysis of the distillate has shown it to contain acetone, methyl ethyl ketone, diethyl ketone, pyrrole, 2-methylpyrrole, phenol, *o*-cresol, 3,5-dimethylphenol, 1-naphthol, 4-picoline, 2,5-lutidine, 2,6-lutidine, 3,4-lutidine, 3,5-lutidine, 2,3,5-trimethylpyridine, and 3-ethylpyridine. The presence of saturated and unsaturated hydrocarbons and absence of aromatic hydrocarbons have also been indicated.

Early transportation was by horse or mule, and somewhat later the 60-mi narrow-gage Uintah Railroad was constructed. In recent years this railroad was abandoned and sacked gilsonite has been moved by trucks.

A major development in the large-scale utilization of gilsonite took place in 1957 when the American Gilsonite Co. began operation of a $13,000,000 refinery for gilsonite near Grand Junction, Colo.

To supply material for the refinery, gilsonite is mined underground hydraulically with a jet of water and pumped (2000 lb/sq in.) to the surface. Rather than reactivate the old Uintah Railroad or truck the product 182 mi, the gilsonite is transported [illegible] as an aqueo[illegible] water. [illegible] has a [illegible] pass, and has a capacity of 700 tons per day of −8 mesh gilsonite. On arrival at the refinery the gilsonite is dewatered by filtration, melted, and re-

fined by conventional techniques used in the petroleum industry. The only major products are coke for use in the aluminum industry, gasoline for local consumption, and gas, valuable as a fuel. The rated capacity of the refinery is 1300 bbl of gasoline and 250–275 tons of coke per day. Proven reserves of gilsonite available for processing are equivalent to more than 100,000,000 bbl of crude oil. *See* PETROLEUM PROCESSING. [I.A.B.]

Bibliography: H. Abraham, *Asphalts and Allied Substances*, 5th ed., vol. 1, 1945; R. K. Bond, Designing the gilsonite pipeline, *Chem. Eng.*, 64(10): 249–254, 1957; Gilsonite yields coke plus gasoline, *Chem. Eng. News*, 35(32):28–29, 1957.

Ginger

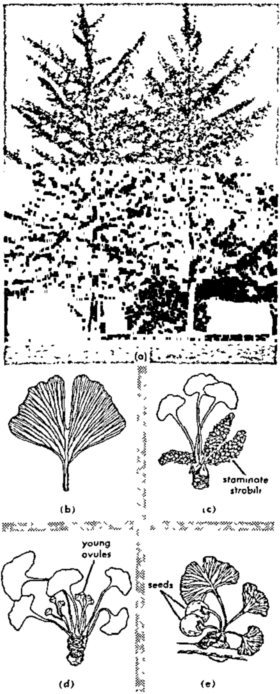
An important spice or condiment. It is obtained from the ginger plant, *Zingiber officinale*, of the ginger family (Zingiberaceae). The plant is a native of southeastern Asia. It is an erect perennial herb having thick, scaly, branched rhizomes which contain starch, gums, an oleoresin (gingerin) responsible for the pungent taste, and an essential oil which imparts the aroma. The rhizomes, dug up

Ginger (*Zingiber officinale*). (USDA)

after the aerial parts have withered, are treated in different ways to produce green ginger or dried ginger. Ginger is used in medicine, in culinary preparations (soups, curries, puddings, pickles, gingerbread, cookies), and for flavoring beverages such as ginger ale and ginger beer. The plant is grown in China, Japan, Sierra Leone, Jamaica, Queensland, and Indonesia. *See* SCITAMINALES; SPICE AND FLAVORING. [P.D.S.]

Ginkgoales

An order of nearly extinct gymnosperms having only one living species, *Ginkgo biloba*, the maidenhair tree. Fossils of several extinct genera have



(a) Two *Ginkgo* trees. (b) *Ginkgo* leaf showing the characteristic dichotomous venation of blade. (c) Spur branch with mature staminate strobili. (d) Spur branch with young ovules. (e) Spur branch with mature seeds. (G. M. Smith et al., *A Textbook of General Botany*, 5th ed., Macmillan, 1953)

been found in the Permian and are abundant in the Jurassic, indicating this was the period of greatest display and widest distribution (*see* GEOLOGY). Toward the end of this period, the Ginkgoales be-

this preserved the species; otherwise, it too would have become extinct.

Ginkgo attains a height of 90 ft and a diameter of 3 ft or more. There are two kinds of branches: long nearly leafless ones, and dwarf branches bearing leaves and strobili. The fan-shaped leaves resemble those of the maidenhair fern, hence the name maidenhair tree. The leaves are dichotomously (forked) veined and often more or less lobed (*biloba*). Male and female strobili occur on separate trees (dioecious). In the male strobilus, microsporophylls are loosely arranged and usually each bears two microsporangia. The female strobilus has a stalk bearing two megasporophylls, each with one ovule. The ovules (usually only one develops) enlarge rapidly. When the ovules are fully developed and become seeds, they have a fleshy outer covering which often causes them to be mistakenly regarded as fruits. As the juicy seed coat disintegrates, it emits an unpleasant odor resembling that of rancid butter. The Ginkgo is widely planted on street borders, home grounds, and in city parks. *See* EMBRYOPHYTA; PLANT KINGDOM. [P.D.S.]

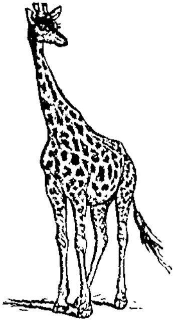
Ginseng

The common name of the genus *Panax*, a group of perennial herbs of the aralia family (Araliaceae), native to the woodlands of the north temperate zone. *Panax schinseng* of Manchuria, extensively cultivated, was in such demand among the Chinese that the supply became insufficient. Then *Panax quinquefolia* of eastern North America was discovered and soon it was being exported to China in large quantities. The price paid for the dried roots was so high that in a relatively short time the collectors nearly exterminated the plants. The Chinese used ginseng as a general panacea for many ills, but there is no evidence that the drug has therapeutic value. *See* UMBELLALES. [P.D.S.]

Ginseng, showing shoot and base. (L. H. Bailey, The Standard Cyclopedia of Horticulture, vol. 3, Macmillan, 1937)

Giraffe

Either of two species of the genus *Giraffa*, both African ruminants of the family Giraffidae, order Artiodactyla. Giraffes have long been favorite circus and zoo animals and are as well known in America as the lion. Giraffes may reach a height of



The giraffe, Giraffa camelopardalis; height 19 ft. (From J. G. Wood, Popular Natural History, Porter & Coates, 1885)

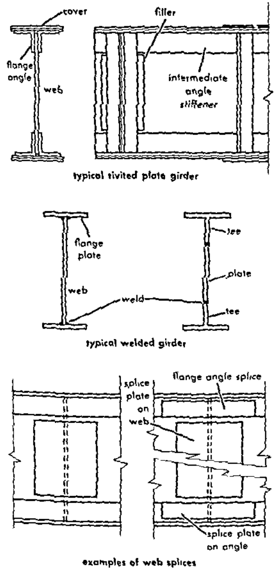
19 ft, but most adults are about 16 ft tall. They browse on the leaves of the mimosa and acacia trees and may go without water for weeks at a time during dry periods, deriving moisture sufficient for survival from their leafy diet. Giraffes travel in small herds of 12–15, led by a large male. *See* ARTIODACTYLA. [J.D.B.]

Girder, plate

A major structural beam, whose sections are built up with plates and angles in the form of an I by riveting or welding. A plate girder is employed primarily to provide greater load-carrying capacity than the deepest available rolled beam. Such girders permit longer spans with greater stiffness than do rolled beams.

Riveted girders consist of a web plate, flange angles, cover plates, end and intermediate stiffener angles, and fillers, connected by rivets as illustrated. Box girders consist of duplicate web and angle assemblies with common cover plates forming a closed section.

Welded girders consist of a web plate welded to flange plates. Flanges may also consist of struc-



Typical riveted and welded plate girders and a girder splice. Rivets not shown.

tural tees welded to an intermediate plate to form the web. Cover plates may be added

Depth of girder sections is selected for economy of web and flange material and is usually one-tenth to one-twelfth the span, with smaller or greater ratios for light or heavy loads. Depth may be controlled by headroom or by architectural considerations.

Webs and flanges. The design of webs is essentially a determination of required thickness for selected depth Two factors are considered: sufficient sectional area to resist maximum shear, and resistance to buckling due to combined bending and shear stresses according to the behavior of an edge-supported plate. Minimum thickness is prescribed to account for possible corrosion. When the ratio of girder depth to web thickness exceeds 70, the shear stress must be reduced; the ratio is never permitted to exceed 170. Reinforcement in the form of stiffener angles or plates is added to avoid web buckling.

Dimensions of the flange acting with the web must furnish the required section modulus according to the flexure formula $S = M/Z$, where S is stress, M is bending moment, and Z is section modulus (*see* LOADING, TRANSVERSE). Flanges in riveted construction consist of angles and plates. The material is used effectively when cover plates provide a large proportion of the flange area and extend only over partial lengths of the span. With only a small contribution by the web, the resisting moment is provided primarily by the flanges. Design proceeds by trial sections determined from the bending moment, assumed effective depth, and allowable stress. The properties of the trial section are computed and the resisting moment compared with that required; further modifications of the section are made until the bending moment meets the requirement. Cutoff points for cover plates depend on variation of moment. Flange dimensions must also be investigated for lateral buckling. Rivets must transfer shear from cover plates to angles and from the entire flange to the web.

Cover plates, attached to the flange angles, form part of the girder flange area. Usually one cover plate extends over the full span length, with additional plates extending over partial lengths in accordance with bending-moment requirements. Cover plates must extend beyond theoretical termination points, permitting enough rivets to develop the plate tension and reduce the stress at the abrupt change of section where stress concentration exists. For welded girders, multiple cover plates are not economical because of additional welding, but the flange may consist of butt-welded plates of variable thickness.

Stiffeners. Riveted angles or outstanding welded plates are attached to the girder web to increase stiffness. Two types of stiffeners are required: end stiffeners over bearings at supports or at application points of concentrated loads, and intermediate stiffeners to prevent web buckling.

Transverse intermediate stiffeners are placed at intervals along the span to increase shear-buckling resistance of the web when the shearing stress exceeds the allowable value for an unstiffened web. Stiffeners must be used when the ratio of web height to thickness exceeds 60. The spacing must provide the necessary web resistance which depends on the ratio of web depth to length between stiffeners Stiffeners must be sufficiently rigid to prevent their buckling with the web.

Bearing stiffeners reinforce the web at reactions and at concentrated loads. Their function is to avoid local buckling and to transmit the load to the web, thus relieving flange rivets of vertical load. Pairs of angles or plates are riveted or welded to the web. They behave as compression members. The outstanding angle legs or the welded plate must have sufficient contact area with the flange, be adequately attached to the web, and be able to transmit the load without themselves buckling.

Stiffeners and part of the web plate function as modified columns.

Longitudinal stiffeners are angles or welded plates attached longitudinally to the girder web. They increase resistance to web buckling caused by bending stresses and are more effective in the compression zone above the neutral axis. Their resistance to axial compression and to bending is influenced by the spacing of transverse stiffeners.

Splices provide continuity of flange plates and angles or web plates composed of more than one piece, necessitated when full lengths are not procurable or when shorter lengths are more readily fabricated. Welded girder flanges are joined by butt welds. Flange angle splices employ short angles, overlapping the cut pieces and attached by the flange rivets. Usually, opposite angle splices are staggered and located where cover plates provide excess section. Welded web splices are butt welds which develop full plate strength. Riveted girders employ plates on each side of the web between flange angles. Flange angles, continuous across part of the cut web, act as an effective splice where excess flange area exists. Plates may be added to the angles. Rivets attaching splice plates resist shear and variable bending forces in the web. [W.J.KR.]

Bibliography: E. H Gaylord and C. N. Gaylord, *Design of Steel Structures*, 1957; C. D. Williams and E. C. Harris, *Structural Design in Metals*, 2d ed., 1957.

Girvanella

A genus of fossil algae. The fossils consist of small pellets or irregular masses of tiny thick-walled tubular filaments. Commonly, the tubes are 8–20 μ in diameter. The genus has a known geologic range extending from the Cambrian to the Cretaceous. It was an important rock builder during the Cambrian, Ordovician, Silurian, Mississippian, and Jurassic. See ALGAE FOSSILS. [J.H.J.]

Bibliography: P. Frémy and L. Dangeard, Sur la position systèmatique des *girvanelles*, *Bull. Soc. linnéenne Normandie*, 8(8):101–112, 1936; A. Wood, The type species of the genus *Girvanella*, *Paleontology*, 1(1):22–28, 1957.

Glacial epoch

An informal popular name which, with its synonym Great Ice Age, refers to the prehistoric time characterized by extensive glaciers and related phenomena, usually supposed to have occupied roughly the last 1,000,000 years of geologic time. It corresponds approximately to the Pleistocene Epoch of stratigraphic literature. The primary characteristics of the glacial epoch were the repeated fluctuation of mean temperatures of air and sea water through a range of several degrees C, and the repeated accumulation of glacier ice on land areas. These events were accompanied by conspicuous fluctuations of sea level, regional subsidence of the earth's crust beneath the extra loads imposed by ice sheets (the largest glaciers), fluctuation of rainfall in wide regions not covered with glacier ice, and extensive migrations of plants and animals. Compared with the general progress of change during geologic time, these events occurred rapidly.

Temperature. The glacial epoch was actually a group of cold glacial episodes alternating with warmer intervals. The record shows evidence of temperatures at times lower than those of today and at other times higher.

Temperatures lower than those of today at the same localities are estimated mainly from five lines of evidence: (1) extent of the former ice sheets, which reached south to latitude 50°N in Europe and to latitude 38°N in North America; (2) difference in altitude of the former climatic snow line on mountains, as approximated by the floors of abandoned cirques, from the snow line of today —measured directly this amounts in places to more than 1000 m; (3) former extent of perennially frozen ground inferred from geologic features, compared with that of today; (4) occurrence of temperature-sensitive fossil plants and animals, both terrestrial and marine, beyond the present ranges of the same or closely related forms; and (5) temperatures of surface sea water, determined directly by $O^{16}:O^{18}$ ratios in the shell substance of pelagic Foraminifera now buried in sea-floor sediments and recovered from core borings in the deep-sea floor. See GEOLOGIC THERMOMETRY.

Equatorial temperatures 2°–5°C lower than those of today are suggested by the criteria indicated in (2), (4), and (5) above. Much greater departures are indicated for continental stations in high latitudes. Temperatures higher than those of today are indicated by (4) and (5). From (5) it appears that in the tropical Atlantic-Caribbean area, Pleistocene temperatures ranged from ~5°C lower to 1°C higher than those of today. At middle-latitude stations, temperatures ~2°C higher than those of today are indicated by (4). Thus the range of temperature throughout the Pleistocene may possibly have been as much as 6°C near the equator and even greater in higher latitudes.

Precipitation. Little is known about the range of precipitation values during the Pleistocene. Evidence of temporary lakes, some of large size, in areas that are dry today, such as western United States, East and North Africa, and Central Asia, indicates increased rainfall and reduced evaporation in those areas, but such evidence does not indicate whether the lakes represent a general increase of precipitation or merely an areal redistribution of an unchanged precipitation total. The belief, once widely held, that increased precipitation values are required to build the extensive Pleistocene glaciers, does not seem necessary, since with sufficiently reduced temperatures, and with time for snow accumulation, glaciers should have been able to grow with snowfall even below today's values. Meteorologic models have assumed both greater and less general precipitation during the glacial times. In southern Africa, at least, ancient sand dunes, ancient soils, and other evidence sug-

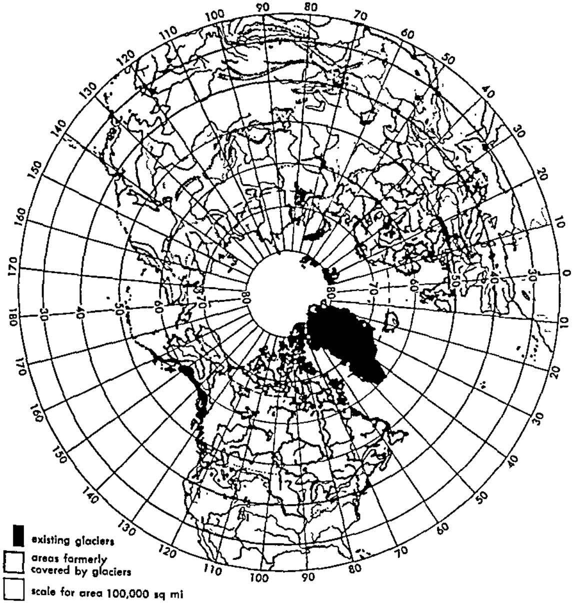


Fig. 1. Glacial map of the North Polar Hemisphere. (Compiled by R. F. Flint from various sources, in Glacial Map of North America, Geol. Soc. Am. Spec. Paper 60, 1945)

gest former climates slightly drier than that of today, but their chronology has not been established.

Fluctuation of climate. In North America four major cold episodes are identified, in Europe four or five. The principal evidence consists of successive bodies of glacial drift and loess, separated by mature ancient soils and by nonglacial sediments containing fossils that imply climates as warm as those of today or warmer. The sediments recovered in cores from beneath the tropical Pacific and Atlantic Oceans show temperature fluctuations, expressed in various ways, that are compatible with the terrestrial data; these may also indicate still earlier fluctuations not yet recognized on the lands. The fluctuations are commonly referred to as glacial ages and interglacial ages, respectively. *See* Marine sediments.

Distribution of glaciers. The extent of the area formerly covered by glaciers is determined mainly from geologic evidence of erosion and deposition by the ice. In general, the central parts of wide glaciated regions have undergone net erosion, having been denuded of their soil cover. Much bare bedrock is exposed at the surface, polished and striated as a result of the passage of ice over it. In contrast, the peripheral parts have received a general mantle of glacial drift averaging a few meters in thickness. Elongate streamline forms fashioned by the ice from bedrock and drift, positions of rock fragments of known origin in the bedrock,

striations, moraines, and eskers afford a basis for determining the directions of movement of the glacier ice. See GLACIATED TERRANE.

Major glaciers, which almost completely covered even the mountainous parts of the terrain on which they lay, included the Antarctic, Greenland, Laurentide (in North America east of the Rockies), and Scandinavian (in northern Europe) ice sheets. Smaller glaciers of various kinds occupied highlands in all latitudes, including the equatorial. The area covered by the recurring Laurentide ice sheet, 13×10^6 km², extended from the Arctic Ocean to a line passing through the vicinities of New York City, Cincinnati, St. Louis, Kansas City, and the Dakotas. When at its maximum, that ice sheet was larger than the existing Antarctic ice sheet. The recurring Scandinavian ice sheet, some 5.5×10^6 km², reached southern Britain, northern Germany, central Poland, and southern European Russia. The margins of the successive glaciers are subparallel, suggesting that the same group of controls operated without major changes during each of the recognized glacial ages

The combined area once covered by glaciers is estimated at 45×10^6 km², about 30% of the present land area of the world. Today glaciers cover nearly 15×10^6 km², about one-third as much area (Fig. 1)

The thickness of former glaciers is very difficult to determine, most of the thickness values quoted are rough estimates and some are hardly more than guesses. Indirect evidence suggests that the ice sheet over eastern Scandinavia may have attained a thickness greater than 2500 m. Direct evidence implies that the Laurentide ice sheet had a minimum thickness of 1600 m over central New England, with the maximum unknown. Meteorologic theory strongly implies that the ice sheets in America and Europe were substantially thicker in their southern and maritime portions, which must have received the most snowfall, than in their northern and continental portions (Fig. 2).

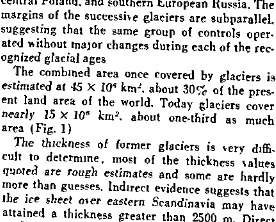


Fig. 2. Diagram illustrating development of an ice sheet such as the Laurentide. Vertical scale much exaggerated. (From R. F. Flint, Glacial and Pleistocene Geology, Wiley, 1957)

The world distribution of former glaciers is reasonable in terms of today's atmospheric circulation system, suggesting that no radical departures from that system need be invoked to explain Pleistocene glaciations. See GLACIER.

Effect on the earth's crust. The coasts of Fennoscandia and northern North America (including the Great Lakes) are marked by a succession of shorelines that have been deformed systematically in such a way that when reconstructed they constitute structural domes whose central parts approximate the central parts of the former ice sheets. The shorelines date from the time of the last deglaciation and shortly thereafter. Tide-gage records show that the upwarping, with little change in form, is still in progress. The crust is regarded as having been depressed isostatically beneath the excess loads created by the ice sheets, with gradual recovery toward its preglacial position as the ice melted. The shorelines at any place were created during the relatively short intervals between deglaciation and crustal recovery. Altitude relations of the shorelines imply that the rate of recovery has been discontinuous, possibly because of variation in the rate of deglaciation and because of temporary expansions of the otherwise shrinking ice sheets. The movement still in progress is occurring at a rate of ~1 mm/(100 km) (year) in the Great Lakes region, and is affecting the depths of Great Lakes harbors. See WARPING, EARTH CRUST.

Sea level changes. Beaches and other marine shorelines of Pleistocene or presumed Pleistocene age occur at various altitudes above sea level, and are detected by echo sounding at depths down to at least 100 m. Submerged stream valleys, deltas, and wind-blown sand have been identified off various coasts. Tide-gage measurements at harbors on the Atlantic and Gulf Coasts of the United States indicate that sea level is rising relative to the land; between 1930 and 1948 the rate was ~6 mm/year. The rise is believed to be an actual rise of sea level, and not a result of sinking of the coast, because of agreement of values at widely separated stations. These data are reasonably explained as an effect of the building and melting of glaciers in response to Pleistocene changes of climate. As temperatures, wind velocities, evaporation, and precipitation vary, equilibrium in the sea water-atmospheric moisture-glacier ice system is established at correspondingly varying positions; as glaciers increase, sea level falls. The current rise of sea level is coincident with world-wide shrinkage of glaciers, established by sampling.

The opinion most widely held is that during glacial maxima, sea level stood about 100 m below its present position; however, still greater estimates have been offered. On fairly good though not indisputable evidence it can be said that the interglacial sea levels were at least 30 m above the present level. Marine strand lines, probably interglacial, exist, but the possibility that they have been deformed casts doubt on the significance of their altitudes. An amplitude of 130 m can therefore be

taken as a probable minimum through which the sea level has fluctuated during Pleistocene time. In consequence of this fluctuation continental shelves emerged widely during glacial ages, restricting the habitats of marine organisms, connecting lands now separated, and affording migration routes for land animals and plants. Lands thus formerly connected include England-Ireland, England-France, Alaska-Siberia, Siberia-Japan, Malaya-Sumatra-Java-Borneo, Australia-New Guinea, Australia-Tasmania,

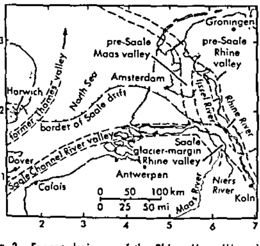


Fig. 3. Former drainage of the Rhine, Maas (Meuse), and Thames rivers as compiled from various sources. (*From R. F. Flint, Glacial and Pleistocene Geology, Wiley, 1957*)

and India-Ceylon. All the separating straits are less than 100 m below present sea level. *See* SEA LEVEL FLUCTUATIONS.

Changes of drainage. The great ice sheets blotted out, at least temporarily, the streams throughout the regions they covered. In some areas the slopes of ice sheets and adjacent land were so related that new rivers formed along the glacier margins, and became so deeply incised that they remained in their new positions after deglaciation. Major examples include the Missouri River through North and South Dakota, the Ohio River from Cincinnati eastward, the Thames River near London, and the lower Rhine-English Channel system, now partly submerged (Fig. 3). Many valleys that existed before glaciation were completely filled with drift by the advancing ice sheets and buried from view. They are being rediscovered by subsurface exploration techniques, and the sand and gravel many of them contain are exploited as sources of water supply.

The Great Lakes basins in the northern United States were developed from major preglacial river valleys, partly by glacial excavation of valley floors and partly by the accumulation of drift to form dams (Fig 4). During deglaciation water was held between the glacier margins on the north and high ground on the south The several successive lake levels recorded by former shorelines were controlled by differential upwarping of the crust and the uncovering of new and lower outlets by deglaciation. The chronology of the lake sequence is

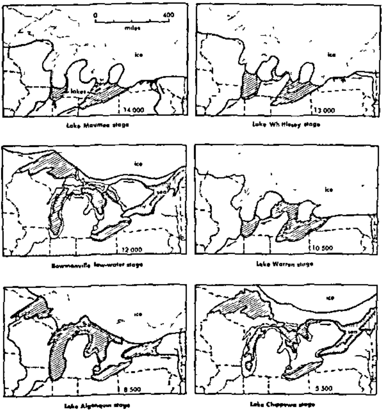


Fig. 4. Evolution of the Great Lakes. Figures indicate approximate number of years before the present time, based on carbon-14 dates. Arrows mark direction of flow along outlets from lakes. (*After R. F. Flint from R. C. Moore, Introduction to Historical Geology, 2d ed., McGraw-Hill, 1958*)

known through the carbon-14 dates of driftwood fragments buried in ancient beaches, of wood buried in contemporary glacial drift, and of contemporary peat. See DATING METHODS.

Glacial drift. The glaciers eroded the unconsolidated mantle and the underlying bedrock down to various depths. Erosion was deepest, amounting in places to more than 1000 m, in the narrow valleys of highland regions. Over wide areas of central Canada and Finland the average depth of glacial erosion is probably no more than a few meters. The eroded material, broken up and comminuted, is spread irregularly over the glaciated regions, particularly the outer parts near the margins of the former glaciers. Some of it was exported by streams draining away from the ice sheets. The Mississippi River carried away from the ice a huge quantity of sediment, part of which still constitutes an outwash fill of gravel, sand, and silt.

Peripheral to the glaciated region are broad blankets of wind-blown silt (loess) picked up mainly from the outwash fills of major river valleys while the discharges of meltwater were in operation. In places more than 100 ft thick, the loess represents a further sorting and redistribution of glacial debris. See LOESS.

Ecologic changes. Study of the accumulated collections of Pleistocene fossil plants and animals has made possible rough reconstructions of ecologic zones in Europe and parts of Asia during the last glacial age. The zones lay far south of their present positions, with tundra and park tundra extending from the margin of the Scandinavian ice sheet southward to the Alps and Carpathians and with forests, so extensive today, confined mainly to the Mediterranean and Black Sea coastal regions. Steppe country lay east of the tundra. Tundra, steppe, and forest each had its appropriate groups of animals.

Zone shifts in North America were less drastic, probably because the arrangement of topography and atmospheric circulation permitted the incursion of warm air from the Atlantic and the Gulf of Mexico [illegible]

[illegible] have consisted mainly of park tundra; forests of spruce and fir lay south of it, and on the west and southwest was steppe, less dry than that of today.

As deglaciation occurred, plants and animals rapidly occupied the region progressively uncovered by the ice, and ecologic zones shifted toward their present boundaries. The rate of shift can be closely inferred from C^{14} dates of lake and bog sediments containing fossil pollen.

Evolution and extinction of organisms. Little evidence regarding the Pleistocene evolution of plants has been accumulated, but evolutionary development, although not striking or conspicuous, occurred among land [illegible] Pleistoc[illegible] possible [illegible] Large mammals adapted to low temperature, such as the woolly mammoth, woolly rhinoceros, and certain musk-oxen, are popularly associated with the glacial ages, but actually they did not appear until late in the Pleistocene. Many of the mammal assemblages were not unlike those found today in comparable climates, with the difference that the ancient assemblages were richer and more varied than those of the present, owing to intervening extinctions. Such extinctions occurred throughout the epoch, but were most marked at a very late time, 5000–20,000 years ago. Opinion is divided as to whether the extinctions were primarily the direct and indirect result of the hunting activity of man, or whether they should be attributed to climatic change or other environmental causes.

Chronology. Although it is frequently said that the Pleistocene Epoch lasted roughly 1,000,000 years, no basis of actual measurement exists as yet. Measurement is confined mainly to C^{14} dating, which thus far hardly extends back of the last 50,000 years. The many hundreds of pertinent C^{14} dates show that the last great glacial invasion of the Great Lakes region began more than 25,000 years ago and reached its maximum extent around 18,000 years ago. From these dates a mean rate of glacier advance through the region of ~160 ft/yr is deduced. Thereafter the ice sheet shrank, with at least two conspicuous reexpansions culminating around 12,000–13,000 and 10,000–11,000 years ago, respectively. The later reexpansion reached close to the sites of Milwaukee, Buffalo, and extreme northern New England. By 8000 years ago the ice sheet had cleared out of all the Great Lakes basins and by 5000 years ago it may have disappeared completely, although evidence on the latter event is scanty. One climatic fluctuation, possibly culminating around 5000 years ago, carried mean annual temperat[illegible] to as much [illegible] last 12,[illegible]ope, as [illegible]lly contemporaneous with those in North America. Although less well established, similar parallelism seems to characterize the entire period of the last 50,000 years as well. C^{14} dates of selected layers in sedimentary cores from ocean floors and from lake sediments in a dry basin in the western United States confirm a cold wet climate changing toward a warmer and drier climate sometime between 11,000 and 18,000 years ago.

Stratigraphy. Designations of glacial and interglacial time units for central North America and for the Alps region in Europe are shown in the table.

Interglacial units are italicized. The Wisconsin and Würm units, about which the most is known by virtue of the excellent state of preservation of the evidence, are further subdivided, but universal agreement on the basis of subdivision is lacking. The time correlation between America and Europe, implied by the chart, is considered probable, but is not by any means established.

Pleistocene time units

Central North America (Glacial) (*Interglacial*)	Alps region, Europe (Glacial) (*Interglacial*)
	Würm
	Riss/Würm
Wisconsin	
Sangamon	Riss
Illinoian	*Mindel/Riss*
Yarmouth	Mindel
Kansan	*Günz/Mindel*
Aftonian	Günz
Nebraskan	*Donau/Günz*
	Donau

Causes of climatic fluctuation. A large number of theories have been advanced to explain the Pleistocene climates, which fluctuated more widely and rapidly than had those of any time within the preceding 200,000,000–250,000,000 years. Although the fossil record suggests that mean atmospheric temperatures in middle north latitudes fell by some 10°C during the 60,000,000 years preceding the Pleistocene, sharp fluctuation seems to have begun with the Pleistocene, as did the glaciation of middle latitudes. There is no evidence that the fluctuation was periodic, and the evidence of recent millenia implies that fluctuation was broadly synchronous throughout the world. Present-day temperatures appear to be close to the interglacial highs.

This and other evidence mentioned earlier suggests an extraterrestrial cause or causes. The cause most frequently appealed to is fluctuation in the amount of energy emitted by the sun; opinions differ concerning the competence of variations in the system of heat transfer within the earth's atmosphere. Alternative theories, appealing to shifts of the earth's crust or of individual continents relative to the poles, to geometric variation in elements of the earth's orbit, to changes in ocean currents, and to variations in the constituents or turbidity of the atmosphere, appear to be less probable from the facts now known. It is clear that most glaciers, including the largest ones, were closely related to highlands, and that prior to the Pleistocene many highlands were lower or nonexistent. Hence, whatever the causes of the variations in climate, Pliocene and Pleistocene mountain-making promoted the development of glaciers in an important way. Once formed, glaciers favored their own self-extension by reflecting solar heat and by setting up anticyclonic conditions in the atmosphere above them. The ultimate solution of the problem of causes is likely to come from the field of meteorology. [R.F.F.]

Bibliography: J. K. Charlesworth, *The Quaternary Era*, 1957; R. F. Flint, *Glacial and Pleistocene Geology*, 1957.

Glaciated terrane

A region distinctive because it once bore great masses of glacial ice which may have been either mountain or mountain and valley type where high ridges and peaks projected above the ice; or continental, ice cap, or plateau type where the ice spread out over a large area, generally of relatively low relief, and concealed almost all the surface. Since mountain valley glaciers in many places coalesce and spread out over adjacent lowlands in piedmont glaciers, the line of division is not absolutely definite and sharp. The diversely complex regional features of glaciated terranes change with the passage of time and vary from place to place, and these conditions are reflected in considerable diversity of scientific views upon the subject. The first part of this article deals with distinguishing characteristics, including regional evidences of prior glaciers and the difficulties encountered in regional identification. A second part outlines scientific understanding of the principal glaciated regions of the world.

Distinguishing characteristics. The regional features characteristic of glaciated terranes and the local patterns associated with prior glaciation are briefly outlined. Modifications of features with the passage of time and other observational difficulties in such regions are then discussed.

Marks of glaciation. That glacial ice extended beyond its present limits or formerly occupied a region now devoid of glaciers is demonstrated by at least a dozen diagnostic features: (1) Scratches (striae) on bedrock and loose rocks which are exactly like those observed in and around existing glaciers. (2) Rock fragments of all sizes, many of them striated, which are unlike adjacent bedrock (hence termed erratics); some of these occur

Fig. 1. A pit cut in bouldery till. Note unstratified and unassorted materials exposed on sides of pit.

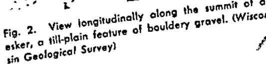


Fig. 2. View longitudinally along the summit of an esker, a till-plain feature of bouldery gravel. (Wisconsin Geological Survey)

higher than their probable bedrock sources and many of them are too large to have been transported by streams. (3) Unstratified and unassorted deposits (till) of clay, sand, gravel, cobblestones, and boulders which were derived in large part by mechanical attrition of bedrock rather than predominantly by chemical weathering. (4) Drainage features unlike those due to normal stream work, including valleys out of harmony with the nature of the bedrock and many waterfalls, rapids, and lakes. (5) Basins completely rimmed with bedrock which are far larger than those due to weathering and lack the regular boundaries of depressions formed by earth movements. (6) Marginal slopes bounding unconsolidated deposits unlike those around the edges of ordinary water-laid deposits. (7) Numerous topographic features which must have been built rather than worn like those due to stream work; these include streamlined hills (drumlins), knobs, ridges, and depressions, all of which can only be constructional (see Figs. 1 to 5) and can be explained only by glacial and glaciofluvial action. (8) Water-laid deposits where such are now impossible because some solid obstruction must have been removed since their formation and there is no indication of erosion to account for this fact. (9) Mountain valleys with an irregular longitudinal profile unrelated to the nature of the bedrock. (10) Mountain valleys which are wider, straighter, and steeper-sided than is normal for the type of bedrock in which they occur. (11) Valley junctions in mountains which are not accordant (hanging) but have falls or rapids. (12) Valley heads and mountainside depressions with abnormally steep sides (cirques). *See* CIRQUE; DRUMLIN; ESKER; MORAINE; TILL.

Observational difficulties. The practical application of the above criteria is fraught with many pitfalls. It is rare to find all or even a majority of them in the same area unless it has been glaciated very recently. As the time since glaciation increases, the criteria become more and more difficult to discover. Common errors in application result from (1) confusion of striae with scratches due to faulting (slickensides), particularly on loose rocks, (2) failure to find striae which are quickly obliterated by weathering and require a cover of till for their preservation for any appreciable fraction of geologic time and consequently are rare on exposed rock ledges, (3) difficulty in demonstrating that some loose rocks are erratics, (4) misinterpretation of deposits of landslides and creep which are deceptively like till, (5) resemblance to drumlins of some rounded hills due to weathering and erosion of massive rock, (6) presence of hanging valleys due to other agencies than glacial erosion, (7) occurrence of erratics transported by icebergs in standing water, (8) removal of glacial deposits by mass movement on hillsides. It is small wonder that opinions concerning the former existence of glaciation in certain localities differ with the personal views of the geologist who observes the evidence.

Principal glaciated areas. Distinctive glaciated terranes are found in every continent beyond the margins of existing glaciers or in regions in which there is no glacial ice at the present time.

North America. Much work on glacial geology has been done in North America. Maps, many of them drawn from air photographs, are available even for some wilderness areas. Continental ice survives in Greenland, and to a lesser extent, in some of the Arctic islands. The Canadian shield of crystalline bedrock is largely scoured rock with

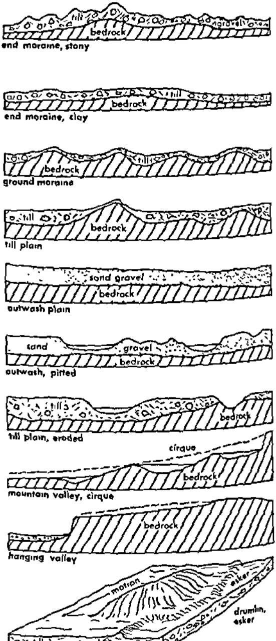
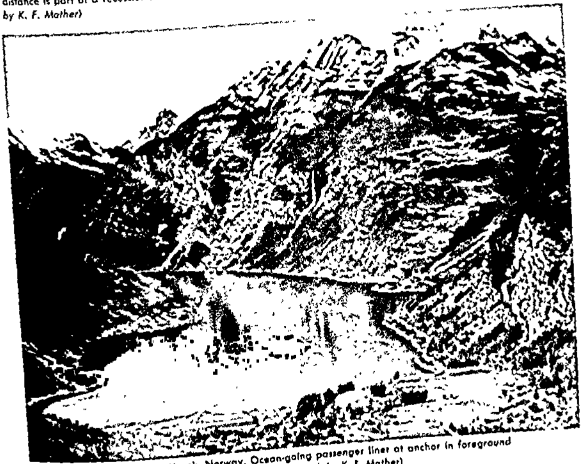


Fig. 3. Cross section profile and block views of typical features of glaciated terrane which are distinctly different from those developed by a normal fluvial cycle of erosion.

[illegible] of submergence. [illegible] stood as low hills [illegible] ong island in the [illegible] on a maturely dissected land surface before late [illegible] distance is part of a recessional moraine left by the retreating ice of the Wurm stage. (*Photograph by K. F. Mather*)



Head of Gieranger Fiord at Merok, Norway. Ocean-going passenger liner at anchor in foreground is more than a hundred miles from the open sea. (*Photograph by K. F. Mather*)

higher than their probable bedrock sources and many of them are too large to have been transported by streams. (3) Unstratified and unassorted deposits (till) of clay, sand, gravel, cobblestones, and boulders which were derived in large part by mechanical attrition of bedrock rather than predominantly by chemical weathering. (4) Drainage features unlike those due to normal stream work, including valleys out of harmony with the nature of the bedrock and many waterfalls, rapids, and lakes. (5) Basins completely rimmed with bedrock which are far larger than those due to weathering and lack the regular boundaries of depressions formed by earth movements. (6) Marginal slopes bounding unconsolidated deposits unlike those around the edges of ordinary water-laid deposits. (7) Numerous topographic features which must have been built rather than worn like those due to stream work; these include streamlined hills (drumlins), knobs, ridges, and depressions, all of which can only be constructional (see Figs. 1 to 5) and can be explained only by glacial and glaciofluvial action. (8) Water-laid deposits where such are now impossible because some solid obstruction must have been removed since their formation and there is no indication of erosion to account for this fact. (9) Mountain valleys with an irregular longitudinal profile unrelated to the nature of the bedrock. (10) Mountain valleys which are wider, straighter, and steeper-sided than is normal for the type of bedrock in which they occur. (11) Valley junctions in mountains which are not accordant (hanging) but have falls or rapids. (12) Valley heads and mountainside depressions [illegible] CIRQUE; [illegible]

[illegible] ...al application of the above criteria is fraught with many pitfalls. It is rare to find all or even a majority of them in the same area unless it has been glaciated very recently. As the time since *glaciation* increases, the criteria become more and more *difficult to discover*. Common errors in application result from (1) confusion of striae *with scratches* due to faulting (slickensides), particularly on loose rocks, (2) *failure to* find striae which are quickly obliterated by weathering and require a *cover of till* for their preservation for any appreciable fraction of *geologic time* and consequently are rare on exposed rock ledges, (3) difficulty in [illegible] (5) resemblance to drumlins of *some rounded* hills due to weathering and erosion of massive rock, (6) presence of hanging valleys due to other agencies than glacial erosion, (7) occurrence of erratics *transported by* icebergs in standing water, (8) removal of glacial deposits by mass *movement* on hillsides. It is small wonder that opinions concerning the former existence of glaciation in certain localities differ with the personal *views of* the geologist who observes the evidence.

Principal glaciated areas. Distinctive glaciated terranes are found in every continent beyond the margins of existing glaciers or in regions in which there is no glacial ice at the present time.

North America. Much work on glacial geology has been done in North America. Maps, many of them drawn from air photographs, are available even for some wilderness areas. Continental ice survives in Greenland, and to a lesser extent, in some of the Arctic islands. The Canadian shield of crystalline bedrock is largely scoured rock with

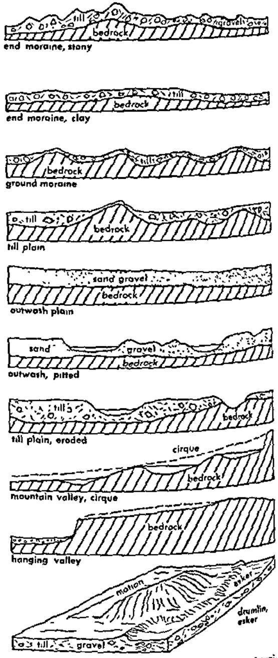


Fig. 3. Cross section profile and block views of typical features of glaciated terrane which are distinctly different from those developed by a normal fluvial cycle of erosion.

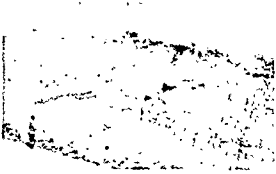


Fig. 4. Cross section along a road cut exposing the till and occasional silt materials (darker materials) in a drumlin.

Fig. 5. A general view over pitted outwash plain. Note lake and lack of well-developed stream and valley system which will probably develop with the passage of time. (Wisconsin Geological Survey)

thin drift, some eskers, and many lakes and extensive areas of lake sediments due to former obstruction of the drainage. In southern Ontario and the prairie provinces, drift is thicker and moraines and drumlins occur. Weathering seems to have destroyed striae on the mountains of Labrador, but it has been thought that this was the point of origin of the continental ice which spread farther toward the west and southwest than in other directions because of the moisture brought by westerly winds. The western mountains of Canada were largely buried by ice, but the spectacular glacial topography is chiefly the result of later local glaciers. In the United States the glacial drift is thick on sedimentary bedrock, particularly in the region south and southwest of the Great Lakes. The bottoms of all the Great Lakes except Lake Erie extend below sea level, and drilling shows that they lie in rock basins eroded in the weaker sediments and evaporites. Plains of lake sediment are found near the shores of the Great Lakes. Glacial drift of several distinct ages has been identified in the Mississippi Valley. The young drift found in the north is termed Wisconsin. In its type locality and throughout much of its extent, its glacial topography is clearly marked with large rugged moraines, pitted outwash plains, eskers, and vast fields of drumlins. Some of the most striking lake districts are in pitted outwash. The lakes there have more rounded outlines than do those in either the thin [illegible]

[illegible]

area of rugged rock topography in which no erratics have been discovered except those evidently brought from outside by lake or stream waters. This region is called the Driftless Area and it has been generally thought that it never was glaciated. The exact boundary of glaciation around this area is definite only on the east side where it is an end moraine of Wisconsin drift. On the other sides the drift thins out gradually, and marginal lakes with ice-rafted boulders make tracing of the border most uncertain. In the nearby glaciated terrane much of the region is strewn with drift too thin to smooth out the rock topography. Although drift might have been removed from hillsides by mass movement, it seems impossible to remove all of it and it is generally recognized that the area escaped glaciation because of its protected position south of the crystalline highland of northern Wisconsin and Michigan.

One of the most spectacular features of glacial origin in Wisconsin is the famous Kettle Interlobate Moraine which records division of the ice into lobes, one in the Green Bay lowland, the other in the basin of Lake Michigan. The Kettle Moraine is the accumulation, chiefly of gravel, in the reentrant angle between these two lobes as their margins melted back.

To the south in Illinois the end moraines become less and less prominent, for the till there has a high content of clay. Ground moraine becomes till plain, partly because of the fluidity of the wet till when deposited, and partly because of buried lake clays. Rock hills rise from these plains like islands from a sea. Postglacial erosion is confined to major stream courses. South of this Wisconsin drift, the margin of which is marked by an end moraine, there is more drift plain, more deeply weathered and more extensively eroded, which is designated as the Illinoian drift. West of the Mississippi River in Iowa the amount of erosion increases abruptly and the Kansan drift occurs west of a definite line. With the same relief, material, and climate the difference can only be due to a longer postglacial time. This conclusion checks with the buried Kansan soil profile. In Iowa, there is a soil profile buried below the Kansan drift; this is the top of the Nebraskan drift, the surface exposure of which, if any, is not definitely recognizable. This region of clay till shows little evidence of moraines. See GLACIAL EPOCH.

The glacial features of Minnesota and the Dakotas are mainly like those of Wisconsin. The clay content of the till increases to the west, making the moraines less evident and outwash rarer. The continental drift meets drift from the Rocky Mountains not far from the foothills. Scenic features

such as glacial cirques, waterfalls, and lakes decrease in the Rockies from north to south. The southernmost glaciation was on the high volcanoes of Mexico.

Owing to the greater moisture available, glacial features are especially prominent in the mountains close to the Pacific. Magnificent fiords extend south from Alaska to near Puget Sound. The sound itself shows deep water but the shores are mainly stratified drift. Some enclosed kettles contain fresh-water lakes. The Yosemite Valley of California is famous for its high waterfalls from hanging valleys, but these are not entirely due to glacial erosion. Such spectacular valleys are few; as pointed out by F. E. Matthes, glacial excavation was profoundly influenced by the amount of fracturing of the otherwise massive rock (Fig. 6).

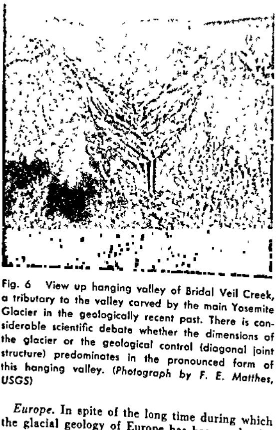


Fig. 6 View up hanging valley of Bridal Veil Creek, a tributary to the valley carved by the main Yosemite Glacier in the geologically recent past. There is considerable scientific debate whether the dimensions of the glacier or the geological control (diagonal joint structure) predominates in the pronounced form of this hanging valley. *(Photograph by F. E. Matthes, USGS)*

Europe. In spite of the long time during which the glacial geology of Europe has been under investigation, few maps of large areas are available. The glacial succession of Günz, Mindel, Riss, and Würm was worked out in the Alps. Mountain centers of accumulation of ice appear to have been of greater relative importance than in North America. Ice came from the Scandinavian mountains, the Urals, and minor centers in the British Isles. The Norwegian coast shows magnificent fiords. Ice spread into the lowlands of Germany and Russia. Three distinct ages have been recognized there.

Asia. Continental ice covered some of northwestern Siberia but its margin is disputed. The dry climate of much of Asia prevented extensive glaciation except in some of the higher mountain ranges.

Other areas. Antarctica is covered by the largest continental glacier now in existence. Local glaciers were or still are present in New Zealand, Australia, and some of the higher Pacific islands. In South America the high peaks of the northern and central Andes and the lower elevations of the southern Andes display glacial features, including fiords along the coast of Chile and extensive moraines in Patagonia. Glaciation is also reported on high peaks in Africa, as well as in New Zealand. *See* FIORD. [F.T.T.]

Bibliography: E. Antevs, Maps of the Pleistocene glaciations, *Bull. Geol. Soc. Am.*, 40:631-720, 1929; R. F. Flint et al. (eds.), *Glacial Map of North America*, Geol. Soc. Am. Spec. Paper 60, 1945; R. F. Flint, *Glacial Geology and the Pleistocene Epoch*, 1947; R. F. Flint, *Glacial and Pleistocene Geology*, 1957; F. T. Thwaites, *Outline of Glacial Geology*, 1956.

Glacier

A naturally accumulating mass of ice that moves in the process of discharging from head or center to marginal dissipation. Glaciers are nourished in an area of snow accumulation which lies above the mean climatological snow line or névé line (*see* SNOW LINE). The most active glaciers are generally found in regions receiving the heaviest snowfall, such as the maritime flanks of high coastal mountain ranges. Exemplifying this are the great westerly facing glaciers of Mount Saint Elias and the Saint Elias Range in south coastal Alaska which are nourished by the heavy precipitation brought in by warm cyclonic air masses moving across the warm waters of the Gulf of Alaska. Similarly, in New Zealand, the vigorous glaciers of the southern Alps lie in the storm tracks of the prevailing westerly wind which brings much greater accumulation to the western than to the eastern slopes. Other such maritime glacial bodies are the Patagonian ice field in the southern Andes; the small ice sheet and ice caps of Iceland; the mountain ice fields of Norway and the Kebnekaise region of Sweden; and the mountain glaciers of eastern Siberia. There are many other glaciers in the high mountains of the middle and equatorial latitudes (as in Peru and East Africa); however, 96% of the world's glacial ice is represented by the vast continental ice sheets of Antarctica and Greenland. Together, these regions contain at least 5,600,000 mi² (14,300,000 km²) of ice cover, or nearly 10% of the total land area of the globe.

The total volume of the world's glaciers, ice fields, and ice sheets can only be estimated, but is at least 24,000,000 km³. This mass of frozen water, if melted and returned to the sea, would raise the sea level 160–200 ft (±60 m). During the Ice Age this volume was probably 300–400% greater,

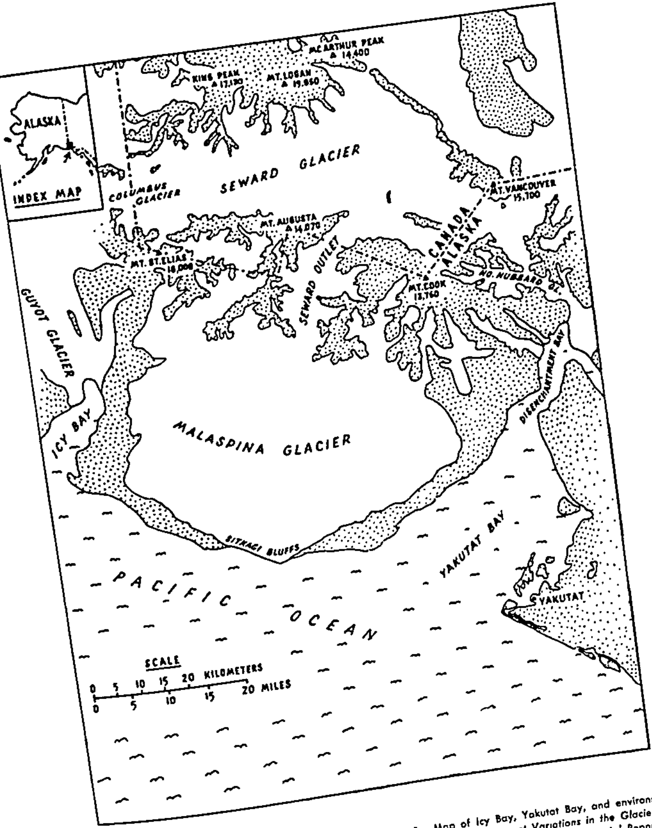


Fig. 1. Map of Icy Bay, Yakutat Bay, and environs. (From M. M. Miller, Recent Variations in the Glaciers of Icy Bay and Yakutat Bay, Alaska, Special Report, Foundation for Glacier Research, 1960)

it was estimated by R. F. Flint in 1957. See GLACIAL EPOCH; TERRESTRIAL FROZEN WATER.

Glaciation and deglaciation. In many regions, glacier activity has formed a continuous series of events throughout the Pleistocene Ice Age and in post-Glacial (post-Ice Age) and modern times. In order to simplify the terminology of past and present glaciation, regions which are presently glaciated are sometimes referred to as glacierized, glacier-covered, or ice-covered. These terms are used instead of the ambiguous and unqualified glaciated for reference to all areas, formerly or currently ice covered.

A glacier at any one time presents only partial patterns of its long-term regime. The study of glacial regimen, or processes of growth and decay, is being actively pursued in glacial regions throughout the world today. Other glaciological studies embrace research in many allied disciplines such as geomorphology, meteorology, and climatology, physics of deformation and thermodynamics, survey and mapping, and pollen and plant ecology.

Materials. Glaciers are composed of three substances: snow, firn, and ice. The main material of glaciers is bubbly glacier ice, of a specific gravity approximating 0.88–0.90. The glacier ice variably contains air pockets and entrapped water bubbles. Because of the proportion of bubbly glacier ice, a mean bulk specific gravity may be taken as 0.90 as opposed to a density of 0.917 g/cm^3 in solid unaerated ice. Below the late-summer glacier snow line (névé line) only glacier ice is exposed. Above the névé line, the other categories of snow and firn (and firn ice) exist to depths of a few feet to hundreds of feet. (Firn is a consolidated granular transition of snow, not yet changed to glacier ice. The process of transformation of snow to firn is termed firnification.) The density gradation between firn and ice is usually asymptotic, but with a relatively sharp line of demarcation between a seasonal snow pack and underlying firn. Arbitrary and approximate specific gravities of new snow are 0.1–0.3; old snow, 0.3–0.5; firn, 0.5–0.75; and firn ice, 0.75–0.88. Firn ice is not considered to be a separate stage of metamorphosis but rather the result of a mixture of partially altered firn and bubbly glacier ice. The processes by which new snow is transformed into dense glacier ice are complex and varied. In middle-latitude regions melt water, compaction, and flow deformation play substantial roles in the metamorphosis. In polar regions, wind packing, mechanical compaction, and flow recrystallization are the prime factors producing glacier ice.

Morphological categories. Glaciers develop numerous forms. Basically, they are of the mountain (alpine) type and of the plateau (polar) type. Mountain glaciers are moderate to small in size and include valley glaciers (ice streams), cirque glaciers, basin glaciers, cliff glaciers, and glacierets. Such terms are self-descriptive and are for the most part associated with varied relief. The valley (ice stream) types are most commonly seen. They serve as outlet glaciers from ice fields such as those found intermittently along the mountainous coast of Alaska and British Columbia and in the Alps, the Caucasus, and the Himalayas. The longest valley glacier in the temperate regions is the Hubbard Glacier with a length of 95 miles in the Alaska-Yukon area. Plateau-type glaciers are of the ice-sheet and ice-cap form, characterized usually by vast size and with relatively flattened surfaces of low relief. These are typical of the Greenland and Antarctic ice sheets, and are sometimes termed inland or continental ice. Intermediate between valley glaciers and ice sheets are piedmont glaciers, which occupy broad lowlands bordering a glacial highland. The best-known example of this category is the Malaspina Glacier in Alaska with an area of approximately 1400 mi^2 (see Figs. 1 and 3).

Geophysical types. Glaciers are classified geophysically into two major groups, polar and temperate. The temperature of a polar glacier is perennially subfreezing, except for a shallow surface zone which may be warmed for a few weeks of each year by seasonal atmospheric variations. In temperate glaciers, the temperature below a recurring winter chill layer is always at the pressure melting point. Because these terms are thermodynamic in meaning but geographical in connotation, it should be pointed out that glaciers of the geophysically polar type can exist at relatively low latitudes and that geophysically temperate glaciers are found at latitudes to the north of the Arctic Circle.

Two subordinate or transitional categories may also be recognized. These are termed subpolar and subtemperate glaciers. In subpolar glaciers the penetration of seasonal warmth is restricted to a

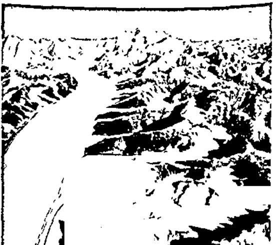


Fig. 2. Snow fields and a mountain-valley glacier system on the southward slopes of the Alaska Range. Note the cirque-head tributaries feeding the main glacier. (MATS, U.S. Air Force)

relatively shallow surface layer, but is greater than in polar glaciers. The transitional subtemperate type of glacier is characterized by a relatively deep zone of annual warming. These subordinate terms are useful because the former may refer to geophysically transitional glaciers of dominantly polar character which still have certain temperate characteristics, and the latter may be used to describe glaciers which are dominantly temperate but which have a tendency towards polar characteristics. The significance of such differentiation is brought out by the close control which ice temperatures exert on flow deformation and glacial fluctuations. Each of the major geophysical categories, polar and temperate ice, can also apply in any one glacier system if there is sufficient range in latitude or altitude for the requisite climatological factors to pertain. Some present-day ice fields and ice sheets which are thermally temperate in some sectors and grade through to geophysical conditions which are thermally polar in another are probably prototypical of the combined geophysical character of the continental ice sheets during the Pleistocene.

Regime. Upon the regime or what may be termed the annual state of health of a glacier system depends its eventual growth or decay. Basically the controlling climatological factors are accumulation (gain) and ablation (loss) over the whole system. The critical zone in this regard is the névé line, the zone which separates the névé area or accumulator from the bare-ice area or dissipator. The height of the névé line varies greatly according to the regional climatological situation and local conditions. On the Juneau ice field in southeastern Alaska this critical line varied over a recent 10-year period roughly between 2600 and 3800 ft (mean 3200 ft). In high-latitude polar regions, the névé line is at or close to sea level. On alpine glacier systems a higher than average névé line usually means a tendency towards a negative regime (more loss than gain), whereas a lower than average névé line is associated with an increase in the accumulation area and a positive regime. In alpine-type glaciers the area ratio of accumulator to dissipator is on the order of 4:1, whereas in the Greenland ice sheet it is roughly 20:1.

In the accumulation zone, there is an increase of retained accumulation in any one year with an increase in elevation until the height of maximum snowfall is attained. This is not always at the highest elevation of land, because it is controlled by the mean freezing level. The greatest snowfall occurs at or near 0°C (32°F). At higher elevations there is commonly less snowfall, the conditions generally being too cold for much snow. The rise and fall of the névé line over a period of years generally parallels the vertical shifts in the zone of maximum accumulation. This vertical shift may be measured by means of soundings, borings, test pit studies, and observations on crevasse walls. In a healthy glacier, the average level of maximum snowfall lies at the elevation of greatest glacier area. A glacier in a poor state of health has usually experienced a rise in its level of maximum snowfall to a height above the level of the main névé area. Also to be considered as a part of the net accumulation of a glacier system is that portion of the summer melt water from the névé in geophysically subpolar or polar (cold) sectors which refreezes as it percolates to depth, becoming a net increment of positive accumulation. In fully temperate glaciers, melt-water percolation drains away almost entirely in subglacial drainage channels and streams. Eventually, this comes out at the snout of the glacier as an increment of net ablation loss.

Structure and movement. A glacier is an ideal field laboratory for the structural geologist, because there are many structures in ice which are comparable to those in sedimentary and metamorphic rocks. A few important structures are primary stratification (that is, bedding strata), secondary fracture structures, tectonic foliation, ablation surfaces, overthrust surfaces, faults, folded structures, and a varied group of deformed sedimentary and structural "bands." There are subsurface transverse ice structures which result from refreezing of melt water in firn. Others manifested at the surface are ice columns and ice veins. Tension and shear fractures such as crevasses and Bergschrunds (crevasses on slopes, exhibiting an overhanging upper lip) are found, as well as moulins (glacier mills, or deep rounded holes caused by water grinding of imbedded stones), cryoconite holes (thermal pits produced by the inmelting of organic or rock fragments), sastrugi (wind-scour features), and others. *All of these are representations of the dynamic processes which affect and control the regime or movement of glaciers.*

Glacier movement is a composite of internal and external movement. The internal movement is domi-

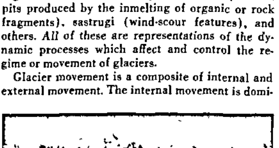


Fig. 3. Aerial oblique view landward over a portion of Malaspina Glacier showing deformed moraines and ice streams. Note glaciated mountain terrane in background. (*MATS, U.S. Air Force*)

nated by a continuous plastic creep (flow deformation) and the external in some cases by a fracture type of discontinuous movement at the bed or near the glacier margins. As a general rule the flow deformation, hence the rate of movement, is greater in the upper portion of a glacier than it is in its basal section (Fig. 4). In glaciers with a healthy and strongly positive regime in the accumulation zone, a strongly dominant proportion of the total mass transfer is the result of actual sliding of the glacier over its bed. Such glaciers are characterized by pluglike velocity profiles in plan view, with the greatest amount of movement expressed in a broad central area (Fig. 5). These glaciers are most effective agents of erosion. In glaciers of equilibrium or negative regime patterns in the névé, the dominant mode of movement appears to be expressed as internal "flow." In plan view the velocity profile is a smooth arched form with associated stream lines. All gradations of movement from the continuous laminar (streaming) type to the discontinuous Block-Schollen plug-type flow may occur in any one glacial system. The proportions depend upon the névé regime pattern, the internal temperature characteristic of the ice, and the configuration of the bedrock channel.

The nature of the plastic internal deformation of glaciers may be treated as a problem in the physics of shear. It is expressed by an exponential relationship between the gravitational stress and the deformation per unit time (creep velocity) with the equation taken as

$$\frac{d\gamma}{dt} = K\tau^n$$

where γ is the shear strain, τ is the shear stress in bars, K is a constant for any given temperature, and n is an empirical constant depending in large measure on the physical character of the ice. Also the exponent n probably depends to some degree on the magnitude of the stress, a factor judged by M. M. Miller as probably significant only in very deep or otherwise highly stressed ice under normal glacier conditions.

Glacial erosion and transportation. Like rivers, glaciers have distinct regions of erosion and transportation. For example, in the high cirque basins of alpine glaciers, rock fragments fall from cliffs or slide into Bergschrunds or marginal moats and become entrained in the ice for transport down the valley. These transported fragments are the critical tools of erosion along the margins and bed of the ice. The larger angular rocks scratch the bedrock and form grooves and striae; the finer materials in contact with the bed do the smoothing and polishing. In such areas, after deglaciation, the striae and grooves are not only proof of former glaciation but they also reveal the former character of flow.

Plucking is also an agent of erosion. This process involves the penetration of ice or of rock wedges into subglacial niches, crevices, and joints in the bedrock. As the glacier moves, it plucks off pieces of jointed rock and carries them away as a supplemental agent of abrasion and plucking. Down-valley ends of jointed hummocks in the bedrock are produced in this manner and are known as roches moutonnées. A sequence of such features produces a steplike longitudinal profile on a glacial valley floor, the steps often coinciding with particularly resistant bedrock.

The repeated fluctuations of glaciers in mountain cirques and over the floors of outlet valleys cause the greatest scouring, transportation, and removal of material. This continuous sequence of processes produces wide-strath highland glacial

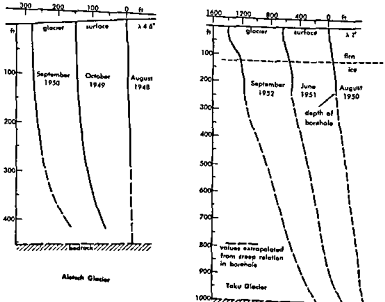


Fig. 4 Vertical profiles of velocity in the Taku Glacier, Alaska, and the Aletsch Glacier, Switzerland.

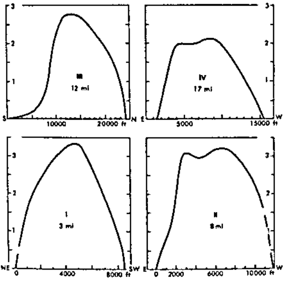


Fig. 5. Partial streaming and Block-Schollen. Types of horizontal surface movement profiles, in the névé-line zone of the Taku Glacier, Alaska. Profiles I and III show partial streaming and profiles II and IV illustrate Block-Schollen. Profiles taken at stated intervals up-glacier from the terminus.

basins and the deep U-shaped outlet valleys which are so common in the Alps, the Cascades, and the Rocky Mountains. Muddy glacier rivers and moraines (lateral, medial, and terminal) are examples of the eroding and transporting power of glaciers. Terminal moraines also testify to continuous glacier activity because they form festooned arrays beyond the ice fronts and express repeated oscillations in recent decades conditioned by the cyclic changes of accumulation and ablation in the distant névés many years before.

Recent and current fluctuation. In post-Glacial time, beginning about 10,000 years ago, the Ice Age was in its waning stage. The result was an almost complete retreat and disappearance of glaciers in the mid-latitudes and tropical regions of the earth. Coincident with this disappearance was the Thermal Maximum (Climatic Optimum) culminating about 5000 years ago. Between 500 B.C. and 1 A.D. a world-wide recrudescence of glaciers took place, associated with a return to harsher climatological conditions. This temperature fluctuation culminated in a colder condition at the beginning of the Christian era. In the fifth and seventh centuries, the evidence suggests that the polar seas were largely free of ice as far north as the pole and that peripheral waters remained rela- [illegible] The records

[illegible]

tury. Two to four centuries ago, there was another world-wide expansion of glaciers and a thickening of polar ice in the new Little Ice Age. Technically the Little Ice Age began with the first resurgence of glaciers about 500 B.C. This latest set of fluctuations is two-fold in nature, having produced a world-wide growth of temperate glaciers reaching their culminations in the mid-eighteenth and late nineteenth centuries. The latest advances on a small percentage of high-level trunk glaciers have continued into the twentieth century, in spite of a general diminution of ice cover around the periphery of some of the lower ice fields. The recent fluctuations in Alaska, Scandinavia, and Patagonia have been manifested in 30–40 moraines over the past 200 years. Such evidence reveals the acute sensitivity of glaciers as indicators of climatic change.

In contrast to the behavior of temperate glaciers, the polar glaciers of Antarctica exhibit a fairly stable regime pattern. This suggests that the significant volume changes in temperate glaciers and the noticeable increase in their internal temperatures have been instrumental factors in accentuating the fluctuation patterns of the Little Ice Age in the middle latitudes. *See* GLACIATED TERRANE.

[M.M.M.]

Bibliography: British Glaciological Society, *Journal of Glaciology*, 1947—; R. F. Flint, *Glacial and Pleistocene Geology*, 1957; M. M. Miller, Phenomena associated with the deformation of a glacier borehole, *Trans. Intern. Union Geodesy Geophys. Gen. Assembly Toronto*, vol. 4, 1958.

Gland

A structure which produces substances essential and vital to the existence of the organism and [illegible]

[illegible] multicellular, tubular, or acinous; (3) according to the manner by which the secretion is delivered to the area of use, as endocrine, when their product is released directly into the blood stream and by this means is carried to the site of use, as exocrine, when the gland delivers its substance to the place of use either directly or by a duct or ducts, and as mixed, when it is both exocrine and endocrine or serous- and mucus-secreting; and (4) according to the manner of cell activity in forming secretion, as holocrine, when the entire cell is consumed in [illegible]

[illegible]

either by modification of cells in situ or as out- growths. In situ glands are either unicellular or the result of cellular modifications, [illegible] sheets of contiguous single cells, or [illegible] groups or nests of cells (Fig. 1). [illegible] glands are multicellular and arise as [illegible] which grow into the surrounding tissue [illegible] incorporate this tissue as part of the [illegible] ture. Outgrowths may be hollow or [illegible] three main types: (1) those glands [illegible] themselves entirely from parental [illegible]

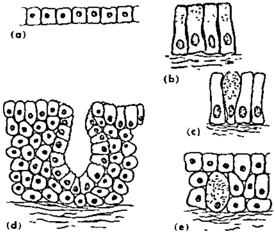


Fig. 1. Intraepithelial gland types. (a) Embryonic epithelium. (b) Epithelial sheet of glandular cells. (c) Unicellular goblet gland. (d) Multicellular gland. (e) Leydig's cell.

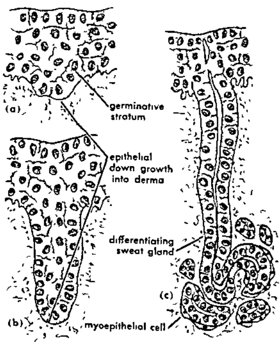


Fig. 2. (a–c) Development of the sweat gland as an outgrowth from the epithelial layer.

as the thyroid; (2) solid, club-shaped masses which become tubelike, and may be simple tubular, simple branched tubular, or compound tubular structures, of which the secreting portion of the gland is tubular, straight, or coiled; and (3) bulbous outgrowths which form sac-shaped secreting units (Fig. 2). Each unit is called an acinus or alveolus. Such glands may be simple acinous, simple branched acinous, and compound acinous. Compound glands, either tubular (Fig. 3) or acinous (Fig. 4), are distinguished from simple glands by the presence of many ducts. The various ducts eventually join one or more common ducts which lead to the surface outlet. Compound glands incorporate surrounding tissues, including blood, lymph, and nerve tissues, into their substance as development proceeds. Such glands become large and have a tendency to subdivide into smaller divisions or lobes. Some compound glands are composed of tubular and acinous structures. They consist of a series of branched tubules which possess saclike, acinous outgrowths from their walls or distal ends.

Secretions. Glands may be classified on the basis of the kinds of secretions they produce. Sebaceous glands secrete oil or oily materials; serous glands, albuminous, watery material; mucous glands, a gelatinous substance; cytogenous glands liberate living cells; and granular glands produce and secrete granules.

Skin glands. All skin glands which include the in situ and epithelial-outgrowth glands are exocrine.

In situ glands are unicellular and hence represent modified single epithelial cells. They remain at the surface of, or within, the epithelial layer of the skin. Small unicellular glands together with larger single-cell glands known as club cells are abundant in the epidermis of fishes. They are mucus-secreting and odoriferous. The glands of Leydig are unicellular structures present in the epidermis of larval urodele Amphibia such as salamanders and newts, and in adult urodeles, such as *Necturus*, which retain a quasi-larval condition in the adult. They are found also in gymnophionan Amphibia. Leydig cells resemble the club cells of fishes. They secrete mucus and some may produce a poisonous substance. Unicellular hatching glands are present in the epidermis of the snout region of

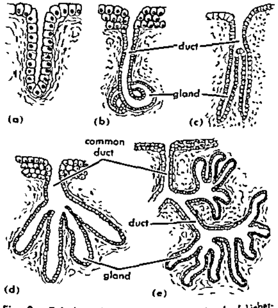


Fig. 3. Tubular glands. (a) Simple (gland of Lieberkühn). (b) Coiled (sweat gland). (c) Simple branched (gastric gland). (d) Simple branched (Brunner's gland). (e) Compound (liver).

frog and toad larvae previous to hatching, and probably also in the larvae of most other Amphibia. Their secretion digests the egg capsule and permits the larva to hatch.

Epithelial-outgrowth glands in their development present bulbous, epithelial outgrowths into the subepithelial tissues, and they may be classified into two groups. One group contains simple, unbranched, or slightly branched acinous glands of fishes and amphibians. In most cases, they are

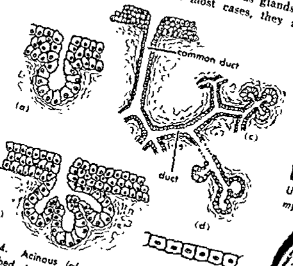


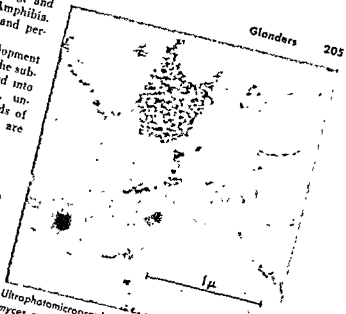
Fig. 4. Acinous (alveolar) glands. (a) Simple. (b) Branched. (c) Compound. (d) Tubuloacinous, that is, the distal secreting units are simple tubes with acinous side chambers.

mucus-secreting although in some fishes and amphibians they are poison-secreting. The second group are simple unbranched, branched acinous, or tubuloacinous glands of reptiles, birds and mammals. The secretion is thick and sebaceous. These glands are abundant in mammals and generally associated with the hair follicles. However, in some areas of the mammalian skin such as the nose, lips, nipples, upper eyelids, around the anus, and external genitals, they arise independently as invaginations of the germinative stratum of the epidermis. See separate articles on the various glands. [O.E.N.]

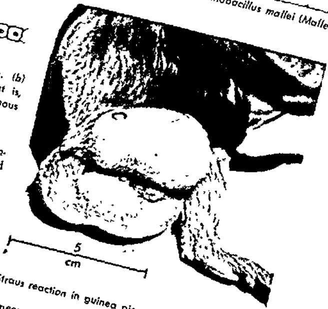
Glanders

A disease of equines, known since the days of Aristotle and Hippocrates, which is occasionally transmitted to man. It is caused by the bacterium *Actinobacillus mallei* (*Malleomyces mallei*). Glanders still exists in horses, mules, donkeys, and man in parts of Europe, Asia, and Africa.

A. mallei is a short, nonmotile, aerobic, gram-negative rod which grows in 48 hours on ordinary media containing glycerin. Characteristic granular staining, antiserum agglutination, and animal inoculation specifically identify the organism. The Straus reaction in hamsters and guinea pigs is an exudative swelling of the scrotum following subcu-



Ultraphotomicrograph of Actinobacillus mallei (Malleomyces mallei)



Straus reaction in guinea pig.

taneous or intraperitoneal inoculation with a suspension of the organism.

Glanders in equines occurs in two forms: (1) acute or chronic respiratory infection, and (2) farcy, a chronic infection of skin and lymphatics. Transmission is through food and water contaminated with nasal or skin discharge. In the conjunctival mallein test a culture filtrate is instilled in the eye, and if the animal has glanders, pus appears in the eye, the eye swells, and in 50–75% of the cases, the temperature rises above 38.5°C. Efficient detection by this test has helped to make glanders almost extinct in highly civilized countries.

Glanders in man is acquired by contact with discharge from infected equines or through accidental laboratory contact. The portal of entry may be respiratory, alimentary, or abraded skin. The disease

is usually acute and runs a stormy course with chills, fever, and marked prostration. Granular pyemic nodules, which become caseous and necrotic, occur in many organs. A milder form of the disease resembling viral pneumonia has occurred in laboratory workers. Diagnosis is established by culture, serologic tests, animal inoculation, and a tuberculinlike skin test with a 1:10,000 dilution of commercial mallein. Combined therapy with sulfadiazine and chloromycetin is effective but must be intensive and prolonged because of the granulomatous lesions and facile bacterial resistance. *See* CHEMOTHERAPY; PNEUMONIA; SKIN TEST. For taxonomy *see* BRUCELLACEAE. [W.R.MI.]

Glass and glass products

Materials made by cooling certain molten materials in such a manner that they do not crystallize, but remain in an amorphous state, their viscosity increasing to such high values that, for all practical purposes, they are solid. Materials having this ability to cool without crystallizing are relatively rare, silica, SiO_2, being the most common example Although glasses can be made without silica, most commercially important glasses are based on this material.

Glass products are enormously varied, including windows, bottles, mirrors, optical instruments, chemical apparatus, building blocks, pipe, doors, cloth, thermal and electrical insulation, electrical condensers, crucibles, chairs, boats, automobile and aircraft bodies, and filters It is even possible to make a glass in which a photographic image can be developed.

Chemical properties. Chemically, most glasses are silicates. Silica by itself makes a good glass (fused silica), but its high melting point (1723°C) and its high viscosity in the liquid state make it difficult to melt and work. For example, the removal of bubbles from silica glass (fining) is very difficult because of its high viscosity. Hence, fused silica products are expensive and are used only when their special properties (low thermal expansion, high softening point, light transmission, and corrosion resistance) are essential.

To lower the melting temperature of silica to a more convenient level, soda, Na_2O, is added in the form of sodium carbonate or nitrate, for example. This has the desired effect, but unfortunately the resulting glass has no chemical durability and is soluble even in water.

To overcome this problem, lime, CaO, is added to the glass, usually in the form of limestone, $CaCO_3$, to form the basic soda-lime-silica glass composition which is used for the bulk of common glass articles such as bottles and sheet (window) glass. Although these are the main ingredients, commercial glass contains other oxides (aluminum and magnesium oxides) and special ingredients to help in oxidizing, fining, or decolorizing the glass batch.

Special kinds of glass have other oxides as major ingredients. For example, boron oxide, B_2O_3, is added to silicate glass to make a low-thermal-expansion glass for chemical glassware which must stand rapid temperature changes, for example, Pyrex brand glass. This type of glass is known as a borosilicate. Also, lead oxide, PbO, is used in optical glass because it gives a high index of refraction. Many other special types of glass have been developed for particular uses, some of which are discussed in this article.

Structure of glass. Physically, glass is an arrangement of atoms quite like that in the liquid state, that is, it has no long-range order. In a crystal, the atoms are arranged in a regular, repeating pattern, but in a glass, although the arrangement in the neighborhood of a single type of atom may be the same throughout the glass (four oxygens around each silicon), the over-all structure lacks periodicity of atomic arrangement (Figs. 1 and 2).

Melting. Production of glass articles begins with batch mixing of raw materials (sand and lime stone) and their melting. For small production and special glass, melting may be done in pots or crucibles containing up to 1 or 2 tons of glass. In large factories, a dozen or so of these pots may be heated by one central furnace into which the pots are inserted and removed through large doors.

Larger batches are melted in large covered furnaces or tanks to which heat is supplied by a flame playing over the glass surface. Usually, these glass tanks are fired regeneratively; that is, the hot exhaust gases pass through an open brick lattice (checker) on one side of the furnace and heat it. After about 15 minutes, the flow is reversed, combustion air (and gas, if used for fuel) is brought in through the hot checkers and preheated, and the exhaust gases go out through checkers on the opposite side of the tank. Most glass tanks are fired by gas or oil; however, auxiliary heating with elec- [illegible] electric [illegible] nk) [illegible] or continuous type. The latter is divided into two

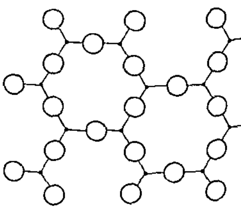


Fig. 1. Schematic representation of the regular arrangement of atoms in a crystal. Large open circles represent oxygen atoms and the small dots silicon atoms. (After W. H. Zachariasen, J. Am. Chem. Soc., 54 3841, 1932)

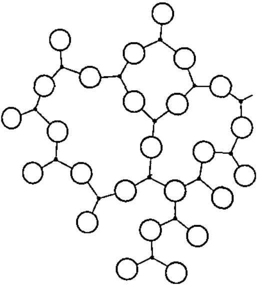


Fig 2. Schematic diagram of the disordered arrangement of atoms in a glass. In this case, as in the case of the crystal, each silicon atom (small dot) is surrounded by the same number of oxygen (large open circles), but long-range order is absent. (*After W. H. Zachariasen, J. Am. Chem. Soc., 54-3841, 1932*)

sections joined by a narrow passage or throat. The raw materials, including any scrap glass (cullet) to be remelted, are charged into one end of the first or melting section. After it has melted, the molten glass, called metal, passes through the throat to the second or refining section, where it stays long enough to be freed of bubbles. From the tank, the glass is taken through forehearths to the forming operation which is different for each type of product.

Heat-treatment. After forming, glass must be slowly cooled or annealed, usually in a long oven called a lehr. The purpose of annealing is to reduce the internal stresses, which can be great enough to crack the glass during cooling. These are created when, because of temperature variations throughout the piece, different parts of the ware become rigid at different times. The opposite of annealing is disannealing or tempering during which the glass is rapidly chilled by a blast of air or even by immersion in liquid. This creates large compressive stresses in the surface of the glass. Because glass invariably fails by tension, this compressive stress must be overcome before the glass will break. Thus tempered, glass is very strong and is used for doors and pipe, but because of the high internal stresses, it is likely to shatter if the surface layer is broken.

Products. The most common glass products are flatware, divided into window glass and plate glass, and container ware such as bottles and jars. Window glass is made by drawing molten glass in a wide sheet from the surface in the forehearth up through the annealing lehr to the cutting machine. Here the continuous sheet, now solidified, is cut into lengths of about 4 ft and the edges, which tend to be thicker than the bulk of the sheet, are trimmed off. The glass is then further cut to size so as to eliminate any blemishes. A variation of the method is to bend the glass to a horizontal path after the initial vertical draw.

Plate glass is formed by passing a continuous sheet horizontally between two rollers at the end of the tank and passing it on through a long, continuous lehr. In the older process, the glass, upon emerging from the lehr, is cut into sheets about 12 ft long and these are mounted in plaster of paris on large flat cars which carry the glass under grinding heads and then under polishers. Generally sand is used for grinding and rouge (iron oxide) for polishing. After one side of the plate is ground and polished, it is turned over, rebedded in plaster and the other side ground and polished. In the newer process, the ribbon of glass remains in one piece until after the grinding, which is done on both sides at once (twin grinding); it is then cut up for polishing. Finally, in either process, the plate glass is inspected and cut into sheets in such a way as to eliminate flaws.

Safety glass is made by bonding two sheets of ⅛-in. plate glass with a transparent organic material, using heat and pressure.

Container ware is largely made on automatic machines; some special pieces may be made by hand blowing into molds, and some very special ware may be made by the free or off-hand blowing of a skilled glass blower. In the automatic process, a stream of glass flows from the forehearth and is

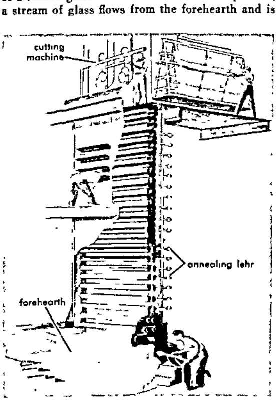


Fig. 3. Drawing window glass in continuous sheets. (*Pittsburgh Plate Glass Co.*)

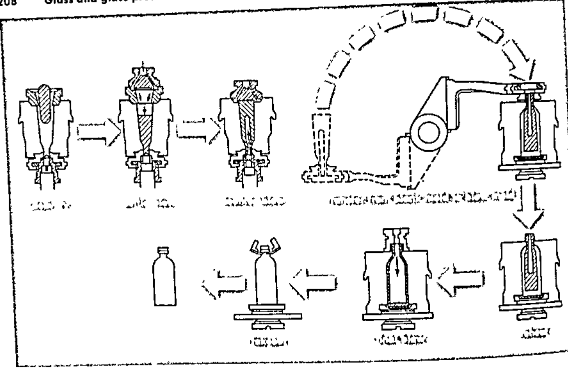


Fig. 4. Automatic bottle manufacture. (Glass Packer)

cut by shears into individual gobs, which are fed to a blank mold. Here the gob is formed into a rough blank, or parison, by either a plunger or compressed air; at this stage the opening of the bottle gets its final shape. The blank mold opens and the parison, supported by the bottle opening, is transferred to the final or blow mold, where it is blown to shape. The bottle is then transferred to a traveling belt which carries it through the lehr to final inspection and packing. There are various modifications of the forming process; in one, for example, the glass is drawn from the tank up into the blank mold by vacuum.

Electric light bulb envelopes are molded on a special machine which converts a fast-moving ribbon of glass into over 10,000 bulbs per minute.

Glass fibers may be either continuous (used for weaving glass cloth) or discontinuous (used for insulation and plastic reinforcement). Continuous fiber is usually made by melting the glass and forming glass balls about the size of marbles; these are then remelted in a platinum-lined tank with small platinum nozzles (spinnerets) in the bottom through which the fiber is drawn. Because fibers have a large surface for a given volume, they are particularly susceptible to chemical corrosion; therefore it is customary to eliminate practically all the alkalies such as Na_2O and K_2O from these glasses. Discontinuous fiber or glass wool is made by blowing a stream of glass into fine filaments by a blast of air or steam. When made from relatively impure natural materials, it is often called rock or mineral wool.

Foam glass, used where a self-supporting insulation is needed, is formed by creating a large volume of small bubbles in the molten glass and allowing it to cool.

Optical glass is generally melted in platinum-lined tanks or pots, from which it is drawn or cast into blanks which can be ground to final shape. An alternate method is to cool the glass in the pot, break it apart, and select flaw-free pieces. Because an important requirement in optical glass is clarity, the purest raw materials are used.

Besides the fairly common types of glass mentioned above, there are hundreds of special types, for example, the milk-white opal glass used for dinnerware and bottles which owes its appearance to tiny crystals grown in the glass.

Special glass. The process of crystallization in a glass, known as devitrification, is generally avoided; however, in some cases it is desired. The new Pyroceram, shaped in the same manner as an ordinary glass, and then, by appropriate heat-treatment, converted to a completely crystalline material, is one example.

Another special glass is sensitive to light, which creates a latent image within the glass, the image being developed by heat-treatment. This glass is used to make signs, and radio and clock faces. A variant of this glass is one in which the portion exposed to light is more soluble than the unexposed portions; it can be used to form, photographically, a fine intricate mesh or screen.

So-called solder glass has a relatively low softening point (below 500°C). It is used to join two pieces of higher-melting glass without softening and deforming them. The solder glass is applied to the joint as a powder and melted to produce a seal, as in a vacuum tube.

Vycor glass is a nearly pure silica glass formed without the production problems of fused silica. The starting glass, a soda borosilicate, is formed and, under heat treatment, separated into two phases, one of which can be leached out. The remaining porous material is nearly pure SiO_2 and can be converted to a dense, clear glass by heating. During this process it shrinks considerably.

Another recent process for producing pure SiO_2 glass involves vaporization of $SiCl_4$ and its hydrolysis to SiO_2.

Properties of glass. The most important properties are viscosity; strength; index of refraction; dispersion; light transmission, both total and as a function of wavelength; corrosion resistance; electrical resistivity; dielectric strength; dielectric constant and loss factor; and chemical durability.

The viscosity of glass is important mainly in its manufacture. The viscosity increases continuously as the temperature drops, the logarithm of the viscosity being nearly proportional to the reciprocal of the temperature. Important points on the viscosity-temperature curve are the melting range (viscosity about 100 poises), the working range (10^{4}–10^{8} poises), the softening point ($10^{7.6}$ poises), the annealing point (10^{13} poises), and the strain point ($10^{14.5}$ poises); the annealing range lies between the last two points.

The strength of glass is of great importance and technical interest. It varies greatly with the conditions of test and is much less than the theoretical values. It is believed that strength really depends on flaws present in the glass surface and that unusually strong glass has fewer flaws or scratches. It appears that small specimens, such as fibers, are much stronger than bulk pieces, but part of the effect may be caused by the rapid chilling fibers receive.

Index of refraction and its variation with wavelength (dispersion) depend mainly on composition In the common types of glass, high refractive index and high dispersion tend to go together, and vice versa. However, in recent years the development of new types of glass has increased the range of properties available to the optical designer.

Light transmission is important for materials such as colored glass and filters. However, it also determines the tint of ordinary glass. The main coloring impurity in glass is iron oxide, which gives a greenish tinge. If a clear white color is desired and total transmission is not important, it can be achieved by adding a decolorizer, which increases the absorption in the blue end of the spectrum, balancing the absorption of the iron. For optical glass, in which maximum transmission is needed, the iron must be eliminated. *See* OPTICAL MATERIALS.

The electrical conductivity of glass, as for most insulators, increases with temperature, the logarithm of the conductivity varying linearly with the reciprocal of the absolute temperature. The surface of glass can be made conducting by a transparent tin oxide coating. *See* DIELECTRICS.

Dimensional stability is of importance in precision instruments, such as clinical thermometers, in which a delayed dimensional change in the glass can destroy the calibration of the instrument. If the glass is not carefully annealed and aged it undergoes a compaction with time because the atoms slowly draw nearer to each other, a condition which is more stable at room temperature. This leads to changes in properties such as density, index of refraction, and strength. [M.C.M.]

Bibliography: G. O. Jones, *Glass*, 1956; G. W. Morey, *Properties of Glass*, 2d ed., 1954; S. R. Scholes, *Modern Glass Practice*, rev. ed., 1952; R. B Sosman, *Properties of Silica*, 1927; J. E. Stanworth, *Physical Properties of Glass*, 1950.

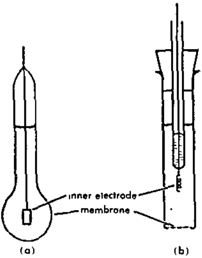
Glass electrode

An electrode or half-cell in which potential measurements are made through a glass membrane. It is used mostly for measurements of the acidity or alkalinity of solutions. The electrode proper, which is immersed in a solution under study, consists of a glass tube having a special glass membrane at one end and containing a sealed-in inner electrode immersed in a suitable solution contained within the glass tube. Connection to the external circuit is made through the inner electrode. Soda-lime glasses give the better performance, and Corning 015 glass having a composition of 72% SiO_2, 22% Na_2O, and 6% CaO is commonly used.

Various types of inner electrodes may be used. These include platinum or gold wire or foil dipped in a buffer solution containing quinhydrone, or a silver chloride electrode, or a calomel electrode in a buffered chloride solution. Buffer solutions are used since the behavior of the electrode depends on the stability of the inner surface of the glass, which in turn depends on the constancy in hydrogen-ion activity. The potential of the inner and outer surfaces of the glass may differ, and give rise to an asymmetry potential which may be determined by placing the electrode in a solution identical in composition to the inner solution. In general, asymmetry potentials are of the order of a few millivolts and may be neglected if the electrode is calibrated with solutions of known hydrogen-ion activity.

The resistance of many glass electrodes is much higher than that of electrodes without membranes; the potential is, therefore, measured with special electrical equipment including electrometers and vacuum-tube amplifiers. Certain types of glass electrodes, however, have sufficiently low resistance so that accuracies of the order of 0.01 in pH may be obtained with common types of potentiometers.

The glass electrode functions as a hydrogen electrode because its potential changes logarithmically in relation to changes in hydrogen-ion activity, except for very low or very high values of hydrogen-ion activity. For very alkaline solutions, the glass electrode is subject to positive errors, that is, it gives too high a slope for potential versus logarithm of hydrogen-ion activity, or gives too a value for hydrogen-ion activity. This



Glass electrodes: (a) Bulb type. (b) Membrane type.

arises from the fact that in very alkaline solutions, the glass electrode tends to respond to metallic ions, notably sodium ions, as well as to hydrogen ions. For use in solutions containing large amounts of sodium ion, special nonsodium glass electrodes are sometimes used.

For very acid solutions, the glass electrode is subject to negative errors, that is, it gives too low a slope for potential versus logarithm of hydrogen-ion activity. This phenomenon arises from the appreciable transport of hydrated protons across the glass membrane. However, in the range of hydrogen-ion activity from 10^{-2} to 10^{-8}, the glass electrode functions well.

The membrane of the glass electrode may be formed in various ways for convenience in use. It may be of the bulb, pinpoint, flat membrane, spiral, capillary, or other type (see illustration). It may be used repeatedly after careful washing, unless etched, broken, or clogged with plastic or other material. Its great advantage over the hydrogen electrode lies in the fact that it may be used in oxidizing and reducing solutions. *See* ELECTRODE POTENTIAL; HYDROGEN ELECTRODE; HYDROGEN ION; SILVER CHLORIDE ELECTRODE. [W.J.H.]

Glaucoma

A disease of the eye characterized by an increase in the fluid pressure within the eyeball. The name was originally applied because of the silver-gray discoloration of the pupil which frequently marks the acute form of the disease.

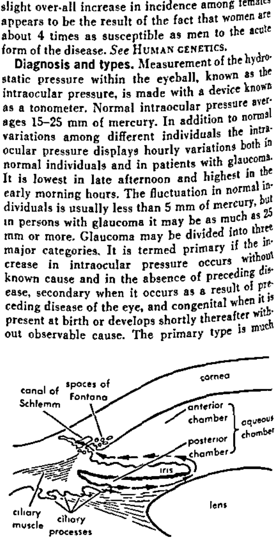
Incidence. Glaucoma is one of the most common and most serious disorders of the eye. It is responsible for 12–15% of the blindness in the United States and is exceeded as a cause of blindness only by cataract. D

the population. Excluding known cases of glaucoma, approximately 1 of every 45 persons over the age of 40 has unrecognized glaucoma in varying degrees of severity. It has been estimated that approximately 10% of such affected persons will ultimately become blind as a result of the disease. This must not be taken to mean that glaucoma carries with it the inevitable prediction that 10% of treated cases will result in blindness, because early recognition and adequate treatment can reduce this figure considerably. This estimate is the result of the unfortunate fact that one of the most common forms of the disease, simple glaucoma, develops slowly and insidiously and is accompanied by few

tion screening tests for persons over 40 years of age have been carried out on a limited scale in a few localities and have been demonstrated to be an effective answer to this problem.

No differences have been demonstrated in racial susceptibility to the disease. Hereditary factors are present in more than 30% of all cases. Over-all figures indicate that the incidence among females is 20–40% greater than that among males. This slight over-all increase in incidence among females appears to be the result of the fact that women are about 4 times as susceptible as men to the acute form of the disease. *See* HUMAN GENETICS.

Diagnosis and types. Measurement of the hydrostatic pressure within the eyeball, known as the intraocular pressure, is made with a device known as a tonometer. Normal intraocular pressure averages 15–25 mm of mercury. In addition to normal variations among different individuals the intraocular pressure displays hourly variations both in normal individuals and in patients with glaucoma. It is lowest in late afternoon and highest in the early morning hours. The fluctuation in normal individuals is usually less than 5 mm of mercury, but in persons with glaucoma it may be as much as 25 mm or more. Glaucoma may be divided into three major categories. It is termed primary if the increase in intraocular pressure occurs without known cause and in the absence of preceding disease, secondary when it occurs as a result of preceding disease of the eye, and congenital when it is present at birth or develops shortly thereafter without observable cause. The primary type is much



Pathway of circulation of aqueous humor in the eye

more common than the congenital and secondary varieties combined, and is responsible for approximately 95% of glaucoma-induced blindness.

Congenital glaucoma. The least common of the three forms, congenital glaucoma, is caused by the absence of the canal of Schlemm, or the vascular channels (see illustration), or both, which provide the drainage pathway for the continuously produced fluid of the aqueous chamber of the eye, the aqueous humor. If treatment is not given (or is not successful) the resultant increased intraocular pressure leads to progressive enlargement of the eyeball, a condition known as buphthalmos. As is the case with most other congenital defects, the cause of the defective canal is not known.

Secondary glaucoma. Blockage of the drainage pathway of the aqueous chamber of the eye can result in secondary glaucoma. Such blockage may be caused by scar tissue formed as a result of injury to or infection of the eye, by the accumulation of protein precipitates or pigment, by bleeding into the eyeball, by a blood clot forming in a large vein of the eye, or by a tumor growing within the eye.

Primary glaucoma. Primary glaucoma may produce symptoms which vary widely from one individual to another and which give rise to a large number of subtypes. These may be broadly classified into two main groups, congestive (inflammatory) and noncongestive (simple).

The congestive form frequently begins with premonitory symptoms (the prodromal stage), consisting of blurred or foggy vision, headaches localized near one or both eyes, and the impression of colored halos around light sources. These are transitory attacks which may continue for years or may culminate within weeks in an attack of acute congestive glaucoma, characterized by severe pain in the eyes, headaches, nausea, and rapidly failing vision which can progress to total blindness within a few days if untreated. More commonly, recovery with only partial loss of vision occurs. This is followed by the stage of chronic glaucoma consisting of recurring attacks, usually less intense and more gradual in onset, each accompanied by additional irreversible loss of vision. Treatment begun at any stage of the disease can only prevent further loss of vision and can never result in recovery of vision already lost.

Simple glaucoma is more common than the congestive type and is more insidious in its development. The symptoms usually consist of nothing more than gradual diminution in vision over a period of years, and are frequently attributed by the patient to increasing age.

The cause of primary glaucoma is not known, but much is known about the factors that influence it and are important in its treatment. The humor filling the aqueous chamber of the eye is constantly being produced and absorbed, resulting in a slow but steady circulation in the eye. The pathway of this circulation is diagrammed in the illustration. The fluid is produced by a process of simple diffusion from the capillaries of the ciliary processes. It then flows out through the pupil over the front surface of the lens and into the cleft formed by the attachment of the iris to the inner surface of the sclera, the iris angle. Here the fluid flows into small channels known as the spaces of Fontana, diffuses through the membrane lining the canal of Schlemm, enters the canal and flows into the venous circulation via the small veins in the outer coat of the eye. An increase in the intraocular pressure can be caused by increased production of fluid by the ciliary process, by interference with the normal path of flow, or by blockage of the spaces of Fontana or the canal of Schlemm preventing normal absorption.

The most important factors influencing these three mechanisms appear to be related to abnormalities of the vascular system of the body. Thus, arteriosclerosis may act by diminishing blood flow in the vessels leading from Schlemm's canal, and in this way impede absorption. Acute attacks of glaucoma occur most frequently in individuals whose blood vessels expand and contract in an exaggerated fashion in response to temperature changes, drugs, and emotional upsets. The erratic fluctuations in blood flow in such cases may lead to increased production of fluid, decreased absorption, or both. About 30% of persons with glaucoma have high blood pressure. In these individuals it is thought that the factors responsible for the increased blood pressure are more important than the fact of high blood pressure itself.

Pathogenesis. The pathologic changes occurring in the eye are caused primarily by prolonged elevation of intraocular pressure. The optic nerve, at its point of entrance into the eye, is pushed backward. This produces tension on the nerve fibers of the retina leading to gradual narrowing of the visual field and ending in complete blindness. The pupil becomes widely dilated and incapable of contraction as a result of atrophy and scarring of the iris. The cornea becomes opaque and ulcerated. The lens may develop cataract, but this is usually of little importance to the patient because blindness will ordinarily have occurred by this time. The blind eye is susceptible to recurrent infection which may eventually require removal of the eye. *See* EYE; EYE DISORDERS. [W.R.AD.]

Glauconite

The term glauconite as currently used has a twofold meaning. It is used as both a mineralogic and morphologic term. The mineral glauconite is defined as an illite type of clay mineral. It is dioctahedral with considerable replacement of aluminum by iron and magnesium. Structural substitutions result in a charge deficiency in both the tetrahedral and octahedral sheets, and interlayer cations seem to balance both of these charges. Calcium and sodium as well as potassium are the interlayer cations. A fundamental characteristic of glauconite is that the unit cell is composed of a single silicate layer rather than the double layer of most other dioctahedral micas. *See* CLAY MINERALS; ILLITE.

Glauconite is known to occur in flakes and as pigmentary material. When used in the morphological sense, the term glauconite often refers to small, green, spherical, earthy pellets. Some of these pelletal varieties are composed solely of the mineral described above, others are a mixed-layer association of this mineral and other three-layer structures.

The mineral glauconite is formed during marine diagenesis. Because of the frequent association of glauconite with organic residues, it has been generally concluded that the presence of organic material is necessary for the formation of this mineral. Glauconite forms in a marine environment from a variety of raw materials. It is known to form as an alteration product of biotite mica, when the alteration takes place under reducing conditions. The deposition of sediment must be slow during this mineralogic transformation to allow complete alteration before burial. If burial is accomplished before alteration, the biotite mica persists. *See* Authigenic minerals; Diagenesis.

Relatively shallow water is another requisite for the formation of glauconite. It has been shown that glauconite forms in the shallow sea in agitated waters which are not highly oxygenated. A reducing environment is also probably necessary for the formation of glauconite.

The magnesium content of glauconite is very uniform, and the ratio of ferric to ferrous iron is rather constant. This suggests that a critical content of magnesium and a particular oxidation-reduction potential might be required for this mineral to form. It has been suggested that certain concentrations of sodium and potassium ions are also necessary for glauconite formation, since the ratio of these cations in the interlayer positions is rather distinctive.

In summary, glauconite forms during marine diagenesis, in relatively shallow water, and at times of slow or negative deposition. In addition, peculiar reducing conditions and particular concentrations of alkalies and magnesium seem to be essential for its formation.

Glauconite readily occurs in pellet form and has been identified in both recent and ancient sediments. It is a major component in some "greensand" deposits and has been used commercially for the extraction of potassium from such sources. *See* Marine sediments. [F.M.W.; R.E.CR.]

Glaucophane

The name given to the monoclinic sodi- [illegible] (color change on rotation in plane-polarized light) in thin sections from yellowish green to deep blue. The crystals are prismatic and sometimes fibrous. The mineral [illegible] association with dense minerals such as lawsonite, $CaAl_2Si_2O_7(OH)_2$; jadeite; pumpellyite, $Ca_2Al_3Si_3O_{12}(OH)$; garnet; and less dense minerals such as epidote, mica, chlorite, and albite. The bulk composition of glaucophane schists is not unusual or particularly rich in sodium. Most often the bulk of the sodium of a rock is present in the feldspars but in the glaucophane schists it is present in glaucophane and jadeite. Because of the higher density of the amphiboles and pyroxenes relative to feldspar the role of high pressures (at least locally) is thought by many to be important in the formation of glaucophane. It is also thought that glaucophane and jadeite may be volumetrically important minerals in the lower regions of the earth's crust if sodium is present in the usual amounts. For the glaucophane schist, the mineral associations and the low metamorphic grade of the rocks deny the presence of high temperatures for these rocks. *See* Amphibole; Jadeite. [C.W.D.]

Glazing

The application of finely ground glass, or glass-forming materials, or a mixture of both, to a ceramic body and heating (firing) to a temperature where the glaze melts, forming a coating of glass on the surface of the ware. Glazes are used to strengthen and decorate the ware, to protect against moisture absorption, to give an easily cleaned, sanitary surface, and to hide a poor body color.

Glazes are classified and described by the following characteristics: surface—glossy or matte; optical properties—transparent or opaque; method of preparation—fritted or raw; composition—such as lead, tin, or boron; maturing temperature; and color. Opaque glazes contain small crystals imbedded in the glass, but special glazes in which a few crystals grow to ½ in. or more in size are called crystalline glazes. A vapor glaze is applied to the ware during firing; an example is the salt glaze applied to sewer pipe by throwing salt (NaCl) in the fireboxes at the peak temperature. The salt vaporizes and then condenses on the ware, where it reacts with the hot surface to produce the glaze.

The most important factor in compounding a glaze, after a suitable maturing temperature has been obtained, is the matching of the coefficient of thermal expansion of the glaze and the body on which it is applied. A slightly lower coefficient for the glaze will place it in compression (the desired condition) when the ware cools. The reverse state (glaze in tension because it has a higher coefficient) leads to the formation of fine hairline cracks, a condition known as crazing. *See* Ceramic technology; Frit; Glass and glass products; Metal coatings. [M.C.M.]

Glide-path indicator

An instrument used in poor visibility to guide the airplane in altitude and direction to a landing at an airport. Two narrow radio beams, projected from the airport, locate the path to be followed in

landing. One beam is for altitude, the other for direction. These radio beams are received by antennas on the airplane, are amplified, and are fed to two milliammeters housed in the same case, called the crossed-pointer indicator. Deviation from the proper direction is observed as a deflection off zero to the right or left of the vertical pointer. Deviation from the proper altitude, which decreases with approach to the airport, is indicated as a deflection off zero up or down of the horizontal pointer. To stay on the correct approach, the pilot maneuvers the airplane to keep the pointers zeroed. *See* INSTRUMENT LANDING SYSTEM (ILS). [W.C.B.]

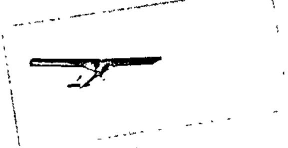
Glider

An unpowered airplane. The glider was used to make the first heavier-than-air flight by man. With the addition of power and further refinements, the airplane was developed. The modern glider has reached a high efficiency and many of its features have been applied to powered aircraft. In the process of development, the word sailplane has come into universal use for high-performance gliders. *See* SAILPLANE.

Types of flight. Motive power may be provided by gravity for simple gliding flight, or by air currents for soaring flight.

Gliding flight is flight from a high point to a lower point. In early gliding, the high point was the summit of a hill or long slope. Early flights were made by accelerating the glider to flying speed by muscular power or by helpers towing the glider into a fairly high wind, and later by the use of elastic cord as a simple catapult. If the gliding angle of the glider was greater than the angle of the slope, a flight somewhat longer than the slope could be achieved.

The next step was soaring flight which occurred when the vertical component of the wind flowing over the slope exceeded the rate of descent or sinking speed of the glider. A gain in altitude could then be made and sustained as long as the condition existed and the glider stayed in the condition. The Wright brothers found that in this way they were able to make flights of up to 10 min in 1911 at Kitty Hawk, N.C. This condition of flight is called slope soaring. From then until about 1930, this was the basic method of soaring flight. Thermal soaring was the next step, accomplished by flying in areas of rising convection currents, which are almost always present to some degree in the atmosphere. The development of thermal soaring and its practice now is principally due to the use of the variometer, calibrated as a sensitive rate-of-climb instrument, which enables the pilot to find the thermal and make the best use of it. By the use of thermal flight, the modern glider is able to fly almost anywhere in the world for extended time and distances over 500 miles in one flight. Other methods of soaring are to make use of clouds and standing-wave phenomena in the atmosphere. These have enabled the sailplane to achieve altitudes of 44,000 ft under the proper conditions.



Two-place training glider.

Glider applications. The training glider is a simple glider with relatively slow flying speed and good stability, enabling the student to learn rapidly and to stay out of difficulty. Training gliders are made for one and for two people as illustrated. They have sufficient performance to enable a competent pilot to soar in reasonably good conditions. Performance is, however, a secondary consideration. The controls are identical to those of a sailplane except that the spoilers or dive brakes are not always used because the performance may not be sufficiently high to require them. The newer training gliders usually incorporate spoilers so that the student may practice their use in glide control. The glider is principally considered a training machine.

The military glider as an assault or cargo vehicle has been largely discontinued by modern air forces since World War II, except for special research applications. Several of the high-speed rocket aircraft, such as the USAF X1, X2, and X15, have been flown as gliders in their preliminary flights. Commercial cargo gliders have no application at present, but there are some aspects of these aircraft that may result in their development in the future. *See* FLIGHT. [E.S.]

Globe, terrestrial

A sphere on the surface of which is a map of the world. The map may be drawn, engraved, or painted directly on the surface but is more commonly prepared as a series of gores, or segments in other designs, to be affixed to the globe ball. Some globes are transparent with a map on the inner surface to prevent soiling or wear. The quality of globe maps varies with the skill of the cartographers, the intended use, and, to some extent, the size or scale. The cost ranges from a few cents for toy models to thousands of dollars for large sizes, depending on the method of construction, the amount of accuracy of detail, and the type of mounting.

Uses. Globes are both artistically interesting and scientifically useful. Their principal value is in stimulating sound concepts of world-wide patterns and in rectifying errors induced by the limitations of flat maps. All flat maps distort the earth's surface patterns, but carefully made globes constitute truer scale models of the earth, with correct areas, shapes, and distances as well as continuity of surface. Globes have long been used as aids in navigation, in the teaching of earth sciences, and as

brary or parlor ornaments. Gradually the field of use has been expanded to include toys and games, advertisements, exhibits, and references for air travel and the behavior of missiles and satellites.

Many modern globes have special attachments to improve their utility. A meridian ring, extending from pole to pole, may be calibrated in degrees to measure latitude. The longitude of points directly beneath that ring will be indicated at the intersection of the ring with the equatorial scale. A horizon ring at right angles to the meridian ring may be calibrated in miles or meters, degrees, and hours to expedite distance and time measurement. A hinged horizon ring may be lifted to serve as a meridian ring, or placed in an oblique position to show great circle routes and distances. Some globes also have a time dial or disk loosely capped over the North Pole. When set to the time of a point directly under the meridian ring, that dial shows the simultaneous time at other longitudes over the earth. In the mid-Pacific the International Date Line is plotted (*see* INTERNATIONAL DATE LINE). Near that line on most globes is a diagram, shaped as a number 8, called the analemma. Its function is to show the latitude at which the sun is directly overhead on each day of the year. With additional attachments globes may

[illegible]

than 16 in. in diameter and most are 12 in. or smaller. Some are as little as 1 in., but one made by Col. Jean-Charles Langlois in 1824 at Paris is 128 ft in diameter. At the Christian Science Publishing House in Boston is a 30-ft translucent ball which people enter to view the world map from the inside. The Babson globe, completed in 1955 at Wellesley, Mass., is 28 ft across with a motor-driven spindle in an outdoor mounting. A 6-ft relief globe was com-

[illegible]

stitute decorative furniture. Others are on tilted spindles or sit as free balls on cradles or inexpensive rings.

Fig 1. Globe gores from the collections of the Library of Congress. (*Instituto Geografico de Agostini, Novara, Italy*)

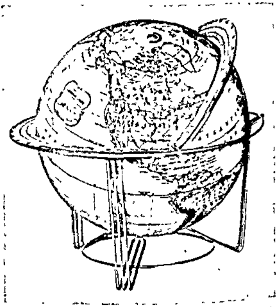


Fig. 2. Globe with hinged horizon ring. Note also in north polar location the small disk or dial which may be separately rotated to indicate simultaneous time all over the globe. (*Rand McNally*)

Famous globes. The oldest terrestrial globe extant was completed in 1492 by Martin Behaim at Nürnberg where the 20-in. original, showing more than 1100 place names, is still in the Germanic Museum. The sphere is of pasteboard and gypsum supported by a wooden frame and covered by parchment. Of almost equal age is an engraved and gilded copper ball, 17 cm in diameter, which was discovered in 1850 at Laon, France. Other globes of the sixteenth and seventeenth centuries were made by such famous European cartographers as Mercator, Hondius, Heyden, Blaeu, Delisle, and Senex, but the most publicized were those made by Coronelli for Louis XIV of France. His largest was 15 ft in diameter, covered with engraved gores (1680), and equipped with a door to permit as many as 30 people to stand inside. Scores of famous globes are collected in the home of the president of the Coronelli-Weltbund der Globusfreunde in Vienna.

Fig. 3. Wilson globe, from collections of the Library of Congress. This globe, dated 1822, is famous as the first made in the Americas. Note example of an analemma, the upper part of which extends north of the horizon ring.

The first globes made in the Americas were constructed by James Wilson at Bradford, Vermont, in 1810, and nearly a dozen globe-making firms in the United States now make a wide variety of globe sizes, types, and mountings. A few make near-globes such as the Icosahedron of 20 spherical triangles or the Dymaxion of six squares and eight triangles, each tangent to a different-sized sphere. [A.C.C.]

Bibliography: I. Fisher and O. M. Miller, *World Maps and Globes*, 1944; E. L. Stevenson, *Terrestrial and Celestial Globes*, 2 vols., 1921.

Globulin

A general name for a group of proteins insoluble in pure water at their isoelectric points, soluble in dilute salt solutions, and precipitated by 50% saturated ammonium sulfate. These properties define the true globulins or euglobulins. Another class of globulins, the pseudoglobulins, is characterized by appreciable solubility in salt-free water but shares the property of the euglobulins in being highly insoluble in 50% saturated ammonium sulfate. Actually no sharp demarcation exists as would be implied by these definitions. Globulins show varying degrees of solubility in water and at various ionic strengths in a spectral fashion, and at some points may actually overlap with some of the albumins in solubility properties. Examples include the gamma globulins of serum (antibody proteins), the lipoproteins of serum (which are divided between the α- and β-globulin fractions), and lactoglobulin, an important component of the milk proteins. *See* GAMMA GLOBULIN; PROTEIN. [D.ST.]

Glow discharge

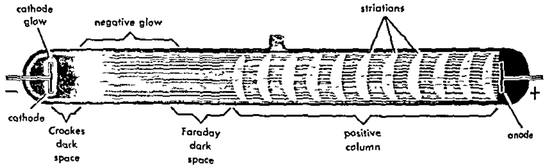
A mode of electrical conduction in gases. Glow discharge commonly occurs under conditions of relatively low pressure and generally in the pressure range of 1–10 mm of mercury. The discharge will typically give off light, so that the region of the discharge appears to glow with considerable intensity. This glow is quite diffuse as contrasted to a higher-pressure discharge, such as a high-pressure arc. Typical currents may be of the order of tens or hundreds of milliamperes, whereas the potential drop may be of the order of 100 volts.

The most important application of the glow discharge is in the so-called voltage regulator or voltage reference tube. This device maintains a relatively constant difference of potential across itself as the current is varied over an appreciable range, and consequently is very useful in cases where a constant reference potential is required.

In terms of the potential-current characteristic, the glow discharge occurs after the potential has been increased so that the Townsend region has been passed. Thus the discharge is field-sustained. On the other hand, a continued increase in current will lead first to the region referred to as the abnormal glow and beyond this to the arc discharge. The transition from the abnormal glow to the arc will generally be almost discontinuous and will be accompanied by a spark. For a discussion of this relationship, *see* ARC DISCHARGE; ELECTRICAL CONDUCTION IN GASES; SPARK, ELECTRIC; TOWNSEND DISCHARGE.

Regions of discharge. There are three main regions of interest in the glow discharge, similar to those in the arc. These are the cathode fall, the positive column, and the anode region. These will be discussed separately, but it is appropriate first to examine some of the general features of the mode (see illustration). The appearance is that of successive more or less well-defined luminous and dark regions. Starting from the cathode, there is a dark space which generally extends for a few millimeters, the Aston dark space. This is followed by a luminous region, also of limited extent, known as the cathode glow. This is succeeded by a somewhat longer dark space designated as the Crookes or Hittorf dark space. After this comes the negative glow region, the boundaries of which are rather poorly defined. Following this is the Faraday dark space, which is also more extensive and poorly defined. This changes gradually into the positive column which is luminous and of length determined by the pressure and distance between electrodes. This region may or may not contain striations, and if present they may be either stationary or moving. At the end of the positive column is a thin layer of greater luminosity designated as the anode glow. Between this and the anode is the anode dark space.

Cathode fall. The events occurring at the cathode are vital to the discharge. The current in the cathode circuit is primarily due to positive ions. However, it is necessary to produce enough electrons



A glow discharge at approximately 0.1 mm pressure. (*From J. B. Hoag and S. A. Korff, Electron and Nuclear Physics, 3d ed., Van Nostrand, 1948*)

at the cathode to maintain the discharge. These electrons will gain energy as they move in the electric field toward the anode, and will produce excitation and ionization. It appears that these electrons are secondary electrons resulting from positive ion bombardment of the cathode surface. The drop in potential which occurs at the cathode will depend on the kind of gas and the cathode material. Generally this potential drop will be a large fraction of the total potential drop across the discharge. The production of secondary electrons by this means is rather inefficient, which explains why [illegible]

[illegible] This can be accomplished only by motion in the electric field, and hence there is a minimum distance which the electrons must move before they can produce excitation and consequent light emission. This explains the existence of the Aston dark space. It might be thought that the cathode glow [illegible]

[illegible] most of this light results from the positive ions that have struck the cathode and are returning to the ground state as neutral atoms. There are two facts of importance in this connection. First the electron density is still rather low at such a short distance from the cathode. Second, the wavelengths present in the radiation indicate transitions involving states of a rather high degree of excitation. These high energy states probably could not be produced by the electrons from the cathode at this point.

The Crookes dark space is actually a region of nearly uniform electric field. Most of the cathode drop occurs in this region, and here the positive ions gain most of their energy before striking the cathode. The electrons from the cathode gain enough energy here to produce cumulative ionization near the end of the space. In the negative glow, which follows, the potential is relatively constant. Here electrons, both from the cathode and from cumulative ionization, lose energy by inelastic collisions and produce a large amount of excitation. The boundary at the anode end of this space is poorly defined owing to the broad distribution in electron energy. The slowing down of the electrons at the end of this region results in a negative space charge. Thus the electrons that move out of this region and into the Faraday dark space gain energy.

Positive column. The beginning of the positive column is the result of excitation by these electrons. The situation in the positive column is the result of a balance between several processes. There is a nearly uniform potential drop which results in ionization throughout this entire region. On the other hand, there must be a loss of ions to make up for this production. This takes place primarily by diffusion to the walls, although there is also recombination. The electrons with their greater mobility will diffuse to the walls, producing a slight negative potential relative to the center of the discharge. This negative potential both limits further electron diffusion and produces positive ion diffusion outward. This process is known as ambipolar diffusion. The positive column is not essential in maintaining the discharge. If the distance between electrodes is changed, with the pressure and current held constant, the extent of the positive column and the potential across the discharge change accordingly. The features of the anode and cathode regions remain unaltered under such a change up to the point where the positive column no longer exists.

Anode region. At the anode end of the positive column, the positive ions are repelled. This produces an increase in electric field, which causes the electrons to gain energy and excite more effectively. Thus the positive column is terminated by a region of increased luminosity called the anode glow.

Other aspects. There are many other aspects of the glow discharge that are interesting and important. One such phenomenon is cathode sputtering. Here the positive ions that are accelerated into the cathode knock out atoms or groups of atoms from the surface (*see* SPUTTERING). Another aspect is that of abnormal glow. The voltage across the discharge remains nearly constant while the current is increased in the normal glow mode. This current increase is accompanied by an increase in the area of the cathode glow. When the glow has completely

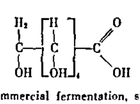
covered the cathode, a further current increase results in an increase in the cathode potential drop, and hence the potential drop across the discharge. This is the abnormal glow. It is characterized by more intense light emission and increased sputtering.

It should be stated that many of the details of the discharge are uncertain. The processes are generally quite complicated. Reliable and accurate measurements are difficult at best, and most of the information is of a qualitative nature. [C.H.MI.]

Bibliography: L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases*, 1939; J. Millman and S. Seely, *Electronics*, 2d ed., 1951.

Gluconic acid

An organic acid with the formula $C_6H_{12}O_7$ produced commercially by fermentation, and used in the pharmaceutical and food industries (*see* FERMENTATION). The compound softens in the temperature range 110–131°C and its structural formula is



In the commercial fermentation, strains of the fungus *Aspergillus niger*, selected for their ability to produce close to theoretical yields of gluconic acid, are employed. Selected bacteria of the genera *Acetobacter* and *Pseudomonas* may also be used *See* PSEUDOMONADACEAE.

The medium consists of inorganic salts and 10–35% of a cheap grade glucose, such as corn sugar. The produced gluconic acid must be kept neutralized, either by continuous addition of sodium hydroxide, to maintain a pH about 6.8, or by initial inclusion of excess calcium carbonate in the medium. Fermentation of sugar concentrations above 15% results in crystallization of calcium gluconate, which retards and eventually stops the fermentation. Crystallization can be prevented at the critical time by addition of small amounts of a complexing agent such as borax.

Fermentor tanks, suited for sterilization under steam pressure and equipped with high speed propellers and a source of sterile compressed air, are used. After sterilization and cooling, the charged fermentor is inoculated with 5–10% of a vegetative (without spores) submerged inoculum grown in a seed tank. The medium is agitated and aerated optimally at the appropriate temperature of 28–35°C. The conversion is exothermic, producing heat, and the fermentors must be continuously cooled. Octadecanol addition controls foaming. The fungus mycelium may be recovered and reused several times, thereby eliminating the growth lag period. Fermentation time is 10–30 hours. The filtered liquor is cooled for crystallization of calcium gluconate. Free gluconic acid syrup (50%) is produced by addition of dilute sulfuric acid and removal of insoluble calcium sulfate. Gluconic acid is used, with sodium gluconate, in solutions used to wash glass bottles. These compounds prevent the precipitation of calcium and magnesium salts which spot bottles. *See* INDUSTRIAL MICROBIOLOGY. [J.W.F.]

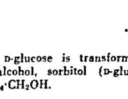
Glucose

A monosaccharide also known as D-glucose, D-glucopyranose, grape sugar, corn sugar, dextrose, and cerelose.

This sugar in free or combined form is not only the most common of the sugars but is probably the most abundant organic compound in nature. It occurs in free state in practically all higher plants. It is found together with D-fructose in considerable concentrations in grapes, figs, and other sweet fruits and in honey. In lesser concentrations, it occurs in the animal body fluids, for example in blood (0.08%) and lymph. Urine of diabetic patients usually contains 3–5%. *See* FRUCTOSE.

Cellulose, starch, and glycogen are composed entirely of glucose units. Glucose is also a major constituent of many oligosaccharides, notably sucrose, and of many glycosides. It is produced commercially from cornstarch by hydrolysis with dilute mineral acid. The commercial glucose so obtained is used largely in the manufacture of confections and in the wine and canning industries. *See* CELLULOSE; FOOD ENGINEERING; GLYCOGEN; STARCH.

Glucose exists in two modifications (α and β). The sugar crystallizes from aqueous solution at temperatures below 50°C as α-D-glucose monohydrate which has a melting point (mp) of 80°C. At temperatures above 50°C but below 115°C, the stable form is anhydrous α-glucose, mp 146°C, $[\alpha]_D$ +113°, mutarotating to +52.2° in water. Above 115°C and below its melting point, the β anomer, mp 148–150°C, $[\alpha]_D$ +19°, mutarotating to +52.2°, is the stable form. β-Glucose can be prepared by crystallization from pyridine or from acetic acid. Ordinary glucose is chiefly the α compound. *See* OPTICAL ACTIVITY.



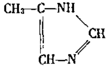
α-D-Glucose

Reactions. Glucose undergoes the general reac-

[illegible]

borohydride, D-glucose is transformed into the hexahydric alcohol, sorbitol (D-glucitol), $CH_2OH\cdot(CHOH)_4\cdot CH_2OH$.

Under the influence of dilute alkalies, it undergoes a series of changes and decompositions which leads to the formation of hydroxy acids, such as lactic acid. When treated with ammonia in the form of ammoniacal zinc hydroxide, glucose is converted, even at ordinary temperature, into methyl iminazole (methylglyoxaline).



Metabolism. D-Glucose is the principal carbohydrate metabolite in animal nutrition; it is utilized by the tissues, and it is absorbed from the alimentary tract in greater amounts than any other monosaccharide. Glucose could serve satisfactorily in meeting at least 50% of the entire energy needs of humans and various animals

Glucose enters the blood stream by absorption from the small intestine. It is carried via the portal vein to the liver, where part is stored as glycogen, the remainder reentering the circulatory system. Another site of glycogen storage is the muscle tissue.

The glucose cycle has been depicted by C. F. and G. T. Cori as follows:

blood lactic acid ↙ ↖
Assimilated glucose → glucose in portal blood → liver glycogen ↘ blood glucose ↗ muscle glycogen

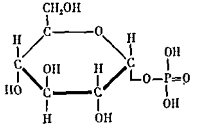
Glucose is readily fermented by yeast, producing ethyl alcohol and carbon dioxide. It is also metabolized by many bacteria, resulting in the formation of various degradation products, such as hydrogen, acetic and butyric acids, butyl alcohol, acetone, and many others. *See* CARBOHYDRATE. [W.Z.H.]

Glucose-1-phosphate

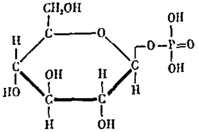
An ester of glucopyranose in which a phosphate group is attached to carbon atom 1 of the sugar, through either the α or β linkage (*see* CARBOHYDRATE). Unlike glucose-6-phosphate, it does not reduce alkaline copper or ferricyanide and is easily hydrolyzed with acid. The biologically important compound, α-D-glucopyranose-1-phosphate, also known as the Cori ester, can be crystallized as the potassium salt, $K_2(C_6H_{11}O_9P) \cdot H_2O$. This compound is an intermediate in the phosphorolytic breakdown and synthesis of the polysaccharides, starch and glycogen, in plants, animals, and microorganisms, and of sucrose in some bacteria. The polysaccharides are synthesized through a reversible reaction, catalyzed by the enzymes known as amylophosphorylases:

$$n \text{ Glucose-1-phosphate} \rightleftharpoons (C_6H_{10}O_5)_n + n\ H_3PO_4$$

α-D-Glucose-1-phosphate is interconvertible with glucose-6-phosphate in the presence of the enzyme



α D-Glucose-1-phosphate



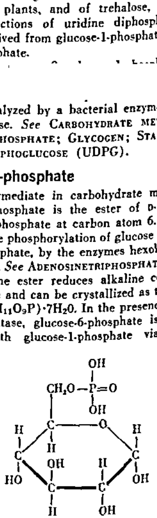
β-D-Glucose-1-phosphate

phosphoglucomutase. The ester is also involved in the metabolism of galactose and in the synthesis of sucrose, by plants, and of trehalose, by yeast, through reactions of uridine diphosphoglucose, which is derived from glucose-1-phosphate and uridine triphosphate.

[illegible] is not [illegible] t is [illegible] alt- [illegible] ose and catalyzed by a bacterial enzyme, maltose phosphorylase. *See* CARBOHYDRATE METABOLISM; GLUCOSE-6-PHOSPHATE; GLYCOGEN; STARCH; URIDINE DIPHOSPHOGLUCOSE (UDPG). [M.D.]

Glucose-6-phosphate

A key intermediate in carbohydrate metabolism. Glucose-6-phosphate is the ester of D-glucopyranose, with phosphate at carbon atom 6. It is produced by the phosphorylation of glucose with adenosinetriphosphate, by the enzymes hexokinase and glucokinase. *See* ADENOSINETRIPHOSPHATE (ATP); ENZYME. The ester reduces alkaline copper and ferricyanide and can be crystallized as the barium salt $Ba(C_6H_{11}O_9P) \cdot 7H_2O$. In the presence of phosphoglucomutase, glucose-6-phosphate is interconvertible with glucose-1-phosphate via glucose-



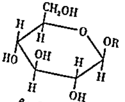
Glucose-6-phosphate

1,6-diphosphate, which acts as a coenzyme (*see* GLUCOSE-1-PHOSPHATE). It is also interconvertible with fructose-6-phosphate through an epimerization, catalyzed by phosphoglucoisomerase. In the presence of glucose-6-phosphate dehydrogenases and a suitable coenzyme, usually triphosphopyridine nucleotide, glucose-6-phosphate is oxidized to the lactone of 6-phosphogluconic acid. *See* BIOCHEMISTRY; CARBOHYDRATE METABOLISM; COENZYME; TRIPHOSPHOPYRIDINE NUCLEOTIDE (TPN).

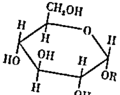
[M.D.]

Glucoside

One of a group of compounds containing the cyclic forms of glucose in which the hydrogen of the hemiacetal hydroxyl has been replaced by an alkyl or aryl group, termed the aglycon group. The difference between α- and β-D-glucosides is only in the configuration of the glucosidic group. Glucosides that contain a six-membered (pyranose) ring are called glucopyranosides. Isomeric glucosides that contain a five-membered (furanose) ring are called glucofuranosides. The latter are relatively easily hydrolyzed and are rare in nature. Glucoside deriv-



β-D-Glucoside



α-D-Glucoside

atives of L-glucose are known also, and the anomers have a mirror-image relationship to the above formulas. Glucosides show the general chemical properties of the glycosides. Although rather stable to alkali, they are hydrolyzed by acid to glucose and the aglycon, which can be an alcohol or phenol. They are also hydrolyzed by specific enzymes called glucosidases, which are found, for example, in almond emulsion, and which may be used to distinguish them from other glycosides.

Naturally occurring glucosides are important in plant metabolism and as drugs, coloring agents, and aromatic principles. The natural glucosides are almost exclusively in the D form, have the β configuration, and are, as a result, levorotatory. The dextrorotatory phillyrin from *Forsythia suspensa*, however, is thought to be an α-D-glucoside. Simple alkyl glucosides are very rare in nature, but methyl β-D-glucoside (aglycon = CH_3) has been isolated from leaves of *Scabiosa succisa*. Most of the biochemical interest in the natural glucosides derives from the properties of the aglycon group and most types of glycosides are represented in the glucosides. *See* GLYCOSIDE.

Many of the anthocyanin pigments of flowers and fruits are glucosides, and often have two glucosidic groups in the 3 and 5 positions of the anthocyanidin nucleus. The coloring matter of saffron (crocin) contains glucosidic groups and represents the newly recognized class of plant sexual hormones.

Saffron is probably involved also in carotene and vitamin A metabolism of plants. Salicin, populin and the vanillin and coumarin glycosides are examples of glucoside derivatives. Many cerebrosides and gangliosides from animal nervous systems also contain the glucosidic linkage, as do the tannins. Cane sugar (sucrose) is a glucoside containing one glucose and one fructose molecule.

Many naturally occurring polysaccharides, such as cellulose, starch, and glycogen, contain glucose units joined together by glucosidic links, and may therefore be regarded chemically as glucosides. *See* MONOSACCHARIDE; POLYSACCHARIDE.

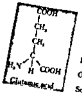
[C.N.R.; R.L.WH.]

Glue

A crude, impure, amber-colored form of commercial gelatin of unknown detailed composition and with an indefinite melting point. Glue does not exist as such in the living organism but is produced by the hydrolysis of animal collagen. It gelatinizes in aqueous solution and dries to form a strong, adhesive layer. The adhesive property of glue is due to the presence of other water-soluble materials because purified gelatin is distinctly inferior as an adhesive. Glues are manufactured by the hot, aqueous extraction of pretreated collagenous material, most commonly animal hides and bones. The first extract contains degraded gelatin and is inferior to the second. Subsequent extractions use hotter water and also produce inferior glue. The glue liquor is concentrated, then dried to a solid of predetermined moisture content. Liquid glue is commonly made from fish collagen because this has little tendency to gel, but it can also be made from animal glue by treatment with acid or certain salts to inhibit gelation. Use of glue is based on the high jelly strength and high tensile strength of the dried film. Applications include use as an adhesive for porous material, chiefly wood, cloth, and paper, a sizing for paper and textiles, and a protective colloid. The term glue is erroneously used to include any material with adhesive properties, such as vegetable glue, casein glue, and phenol-formaldehyde resin glue. *See* ADHESIVE; MUCILAGE.

[E.H.H.]

Glutamic acid



Glutamic acid

Physical constants of the L isomer at 25°C
pK_1 (COOH) 2.19; pK_2 (COOH) 4.25; pK_3 (NH_3^+) 9.67
Isoelectric point 3.22
Optical rotation $[\alpha]_D(H_2O)$ +12.0; $[\alpha]_D$(5 N HCl) +31.8
Solubility (g/100 ml H_2O) 0.84

An amino acid. The amino acids are characterized physically by the following: (1) the pK_1, or the dissociation constant of the various titratable groups; (2) the isoelectric point, or pH at which a dipolar ion does not migrate in an electric field; (3) the optical rotation, or the rotation imparted to a beam of plane-polarized light (frequently the

D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml; (4) solubility. *See* EQUILIBRIUM, IONIC; ISOELECTRIC POINT; OPTICAL ACTIVITY; SPECTROPHOTOMETRIC ANALYSIS.

Glutamic acid forms an alcohol-insoluble calcium salt. The amino acid has many important functions, including the following:

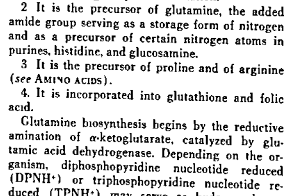
1. It is the principal point by which ammonia enters organic compounds (reductive amination of α-ketoglutarate) and the central distribution point for amino nitrogen by transamination.

2 It is the precursor of glutamine, the added amide group serving as a storage form of nitrogen and as a precursor of certain nitrogen atoms in purines, histidine, and glucosamine.

3 It is the precursor of proline and of arginine (*see* AMINO ACIDS).

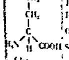
4. It is incorporated into glutathione and folic acid.

Glutamine biosynthesis begins by the reductive amination of α-ketoglutarate, catalyzed by glutamic acid dehydrogenase. Depending on the organism, diphosphopyridine nucleotide reduced ($DPNH^+$) or triphosphopyridine nucleotide reduced ($TPNH^+$) may serve as hydrogen-donor. During metabolic degradation, the most common pathway is through deamination to α-ketoglutarate. A novel pathway exists in the bacterium *Clostridium tetanomorphum*, which ferments glutamic acid to butyric acid, acetic acid, carbon dioxide (CO_2), ammonia (NH_3), and hydrogen (H_2) by a reaction sequence including mesaconic and citramalic acids:



See ARGININE; FOLIC ACID; HISTIDINF; PROLINE. [E.A.AD.]

Glutamine


Glutamine

Physical constants of the L isomer at 25°C
pK_1 (COOH) 2 17 pK_2 (NH_3^+) 9 13
Isoelectric point 5 65
Optical rotation $[\alpha]_D(H_2O)$: +6.3 $[\alpha]_D$(1 N HCl): +31 8
Solubility (g/100 ml H_2O) 3 6 (18°C)

An amino acid. The amino acids are characterized physically by the following: (1) the pK_1, or the dissociation constant of the various titratable groups; (2) the isoelectric point, or pH at which a dipolar ion does not migrate in an electric field; (3) the optical rotation, or the rotation imparted to a beam of plane-polarized light (frequently the D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml; (4) solubility. *See* EQUILIBRIUM, IONIC; ISOELECTRIC POINT; OPTICAL ACTIVITY; SPECTROPHOTOMETRIC ANALYSIS.

Glutamine with its amide group is an important storage form of nitrogen in plants and animals; it also serves as the precursor of certain ring nitrogen atoms in purines and histidine and of the amino group in glucosamine. The biosynthetic pathway for glutamine is from glutamic acid, by the following reaction:

$$\text{Glutamic acid} + \text{ammonia} + \text{ATP} \rightarrow \text{glutamine} + \text{ADP} + \text{phosphoric acid}$$

Adenosinetriphosphate (ATP) is required in the reaction, and adenosinediphosphate (ADP) is one of the products.

During metabolic degradation, glutamine is converted to α-ketoglutarate by either of two pathways: (1) deamidation to glutamic acid, followed by deamination; or (2) transamination of the amide group to a keto-acid acceptor, forming α-ketoglutaramic acid. The latter is then hydrolyzed to α-ketoglutaric acid and ammonia *See* ADENOSINEDIPHOSPHATE (ADP); ADENOSINETRIPHOSPHATE (ATP); AMINO ACIDS; TRANSAMINATION. [E.A.AD.]

Glutelin

A general name for a group of proteins soluble in dilute acid or alkali but insoluble in water and all neutral solvents. Examples are glutenin from wheat and oryzenin from rice. *See* PROTEIN; RICE; WHEAT. [D ST.]

Glycerol

The simplest trihydric alcohol, formula $CH_2OHCHOHCH_2OH$. The name glycerol is preferred for the pure chemical, but the commercial product is usually called glycerin. It is widely distributed in nature in the form of its esters, called glycerides. The glycerides are the principal constituents of the class of natural products known as fats and oils.

Uses. Glycerin is used in nearly every industry With dibasic acids, such as phthalic acid, it reacts to make the important class of products known as alkyd resins which are widely used as coatings and in paints. Because of its valuable emollient and demulcent properties, it is used in innumerable pharmaceutical and cosmetic preparations. It is an ingredient of many tinctures, elixirs, cough medicines, and anesthetics. It is a basic medium for toothpaste.

In foods, it is an important moistening agent for baked goods and is added to candies and icings to

prevent crystallization. It is used as a solvent and carrier for extracts and flavoring agents and as a solvent for food colors.

Because of its humectant properties, it is sprayed on tobacco before it is processed to prevent crumbling and is added to adhesives and glues to keep them from drying too fast. Many specialized lubrication problems have been solved by using glycerin or glycerin mixtures.

Many millions of pounds are used each year to plasticize various materials. As much as 15% is added to cellophane to render it pliable. It is included in meat casings and special types of paper for the same purpose. Sheets and gaskets made from ground cork are plasticized with glycerin.

Of its chemical derivatives, the esters of glycerin are the most important. Nitroglycerin (glyceryl trinitrate) is used in the manufacture of dynamites and propellants. The mono- and diesters of higher fatty acids can be formed by direct reaction of glycerin with the acid or by a transesterification reaction of a glyceride with glycerin. These esters are used as emulsifiers in foods and preparation of baked goods and for modification of alkyd resins.

The production of glycerin has increased at a steady rate. In 1930, 110,000,000 lb was made, and in 1958 the production was 222,000,000 lb.

Properties. When pure, glycerin is a colorless, odorless, viscous liquid with a sweet taste. It is completely soluble in water and alcohol but is only slightly soluble in many common solvents such as ether, ethyl acetate, and dioxane. Glycerin is insoluble in hydrocarbons. It boils at 290°C at atmospheric pressure and melts at 17.9°C. Its specific gravity is 1.262 and its molecular weight is 92.09. It has a very low mammalian toxicity.

Production. Glycerin was first discovered by Carl W. Scheele in 1779 who made it by heating olive oil with litharge. Until after World War II, nearly all the glycerin of commerce was produced as a by-product in the manufacture of soap or from the hydrolysis (splitting) of fats and oils. However, at present, a substantial portion of the material made in the United States is prepared synthetically from propylene.

In the process of soap making, called saponification, fat reacts with aqueous sodium hydroxide. The crude product is coagulated by the addition of salt. The acid portion of the fat combines with the sodium hydroxide to form solid soap, and the glycerin liberated in the reaction remains in the salt solution which is called spent lye.

In the process involving the hydrolysis of fats, they react with water to give the component acids and glycerin. An aqueous solution of the latter is produced, called glycerin sweet water. From this liquid and from spent lye from soap making, glycerin is obtained in much the same way. After a preliminary treatment to remove impurities, the water of solution is evaporated under reduced pressure. The residual glycerin is filtered while hot to remove precipitated salts. For most applications, it is necessary to refine it further by fractional distillation under reduced pressure.

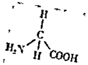
In the synthetic process propylene reacts with chlorine to give allyl chloride. This may be hydrolyzed to allyl alcohol which is then treated with aqueous chlorine to produce a mixture of chlorohydrins. These are then hydrolyzed to glycerin. In an alternative procedure, allyl chloride is treated with chlorine and water to produce a mixture of chlorohydrins, and these are hydrolyzed as in the other method. A dilute solution containing about 5% of glycerin is produced in the hydrolysis reaction in both instances. After the water is evaporated, distillation gives a product with a purity greater than 99%.

Several grades of glycerin are marketed, including high gravity, dynamite, yellow distilled, USP (U.S. Pharmacopoeia) and CP (chemically pure). USP grade is water-white and suitable for use in foods, pharmaceuticals, and cosmetics or for any purpose when the product is designed for human consumption. *See* ALCOHOL; FAT AND OIL, EDIBLE; FAT AND OIL, NONEDIBLE; POLYHYDROXY ALCOHOL.

[B.W.K.]

Bibliography: C. S. Miner and N. N. Dalton (eds.), *Glycerol*, Am. Chem. Soc. Monograph 117, 1953.

Glycine



Glycine

Physical constants of the L isomer at 25°C:
pK_1 (COOH) 2.34, pK_2 (NH_3^+) 9.60
Isoelectric point: 5.97
Solubility (g/100 ml H_2O): 24.99

An amino acid. The amino acids are characterized physically by the following: (1) the pK_1, or the dissociation constant of the various titratable groups; (2) the isoelectric point, or pH at which a dipolar ion does not migrate in an electric field; (3) the optical rotation, or the rotation imparted to a beam of plane-polarized light (frequently the D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml; (4) solubility. *See* EQUILIBRIUM, IONIC; ISOELECTRIC POINT; OPTICAL ACTIVITY; SPECTROPHOTOMETRIC ANALYSIS.

Glycine has many important functions, including the following:

1. It is incorporated intact into purines.
2. The α-carbon and nitrogen atoms are incorporated into the pyrrole rings of porphyrin.
3. It accepts an amidine group from arginine and a methyl group from methionine to form creatine.
4. It is a constituent of the tripeptide coenzyme, glutathionine (γ-glutamylcysteinylglycine).
5. It can accept formaldehyde from hydroxymethyltetrahydrofolic acid to become serine. This reaction is reversible.

The quantitatively most important biosynthetic pathway for glycine is from serine, which transfers its hydroxymethyl group to tetrahydrofolic acid. Smaller amounts arise by the transamination [illegible]oxylic acid, which can be formed by cleavage

of isocitric acid (*see* TRANSAMINATION). Threonine may also be cleaved to glycine and acetaldehyde.

During metabolic degradation, glycine is deaminated to glyoxylic acid, which is oxidized to formate and carbon dioxide. Under some conditions, glyoxylate can be oxidized to oxalic acid. *See* AMINO ACIDS. [E.A.AD.]

Glycogen

The primary reserve polysaccharide of the animal world. It is found in the muscles and in the livers of all higher animals, as well as in the cells of lower animals. Because of its close relationship to starch, it is often called animal starch, although glycogen is found in some lower plants, fungi, yeast, and bacteria. A polysaccharide similar to glycogen was isolated in one case from a higher plant, Golden Bantam sweet corn (*Zea mays*). *See* POLYSACCHARIDE.

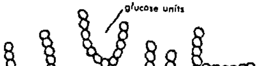
Properties. Glycogen is a nonreducing, white, amorphous polysaccharide which dissolves readily in cold water, forming an opalescent, colloidal solution. It gives a reddish brown color with iodine, is precipitated by alcohol, and has a specific rotation $[\alpha]_D^{20}$ of approximately +200° (*see* OPTICAL ACTIVITY). It is very resistant to the

proteins, and precipitating the glycogen with ethyl alcohol.

The molecular weight of glycogen is usually very high, and it varies with the source and the method of preparation; molecular weights of the order of 1–20 × 10^6 have been reported.

In its biochemical reactions, glycogen is similar to starch. It is attacked by the same plant amylases that attack starch, and like starch, it is degraded to maltose and dextrins. Both glycogen and starch are broken down by animal or plant phosphorylase enzyme in the presence of inorganic phosphate with the production of α-D-glucose-1-phosphate.

Molecular structure. Chemical studies, based on methylation and periodate oxidation procedures, show glycogen to possess a branched structure similar to the amylopectin starch fraction. The molecules of both these polysaccharides consist of chains of D-glucose residues joined by α-1,4 linkages, having similar chains attached through α-1,6 linkages at the branch points (Fig. 1). Depending on the source, the average chain length of a branch (which comprises the average length in glucose units of the outer and inner branches) in a glycogen molecule is 11–18 D-glucopyranose units. In amylopectin, the average chain length of a branch is 22–27.



Diagrammatic representation of the branched structure of glycogen. (After K. H. Meyer)

Synthesis. Two enzymes concerned with the synthesis of glycogen, muscle phosphorylase and branching factor are known to be present in animal tissues. The first enzyme, muscle phosphorylase, in the presence of α-D-glucose-1-phosphate catalyzes the following equilibrium reaction:

$$\alpha\text{-D-Glucose-1-phosphate} \rightleftharpoons \alpha\text{-1,4-linked polysaccharide} + \text{Inorganic phosphate}$$

In this reaction, which requires a trace of glycogen as a primer, linear chains of α-1,4 linked D-glucose units are formed. The second enzyme, branching factor, has the ability to unite such chains through 1,6 linkages, resulting in the formation of a highly branched glycogen structure. It is believed that the combined action of the two enzymes is responsible for the synthesis of glycogen in nature.

The chief carbohydrate of mammalian skeletal muscle is glycogen, there being relatively little free glucose in this tissue. The sequence of reactions in anaerobic glycolysis during muscle contraction begins with glycogen and terminates with lactic acid. C. F. Cori and G. T. Cori have shown that lactic acid diffuses out of muscle into the blood stream, which carries it to the liver where it is converted to glycogen and stored. Liver glycogen is converted into glucose which is secreted into the blood stream. The glycogen of mammalian muscle arises from the glucose carried to it from the liver by the blood. There is, therefore, a close interdependence between carbohydrate metabolism of muscle and liver.

The metabolic formation of glycogen from glucose in the liver is frequently termed glycogenesis. In fasted animals, glycogen formation can be induced by the feeding, not only of materials that can be hydrolyzed to glucose and other monosaccharides, such as fructose, but also of various other materials. A number of L-amino acids, such as alanine, serine, and glutamic acid, upon deamination in the liver, give rise to substances, such as pyruvic acid and α-ketoglutaric acid, that can be converted in the liver to glucose units which are subsequently converted to glycogen. Furthermore, substances such as glycerol derived from fats, dihydroxyacetone, or lactic acid can all be utilized for glycogen synthesis in the liver. Such noncarbohydrate precursors are termed glycogenic compounds. The process of glycogen formation from these precursors is known as glyconeogenesis. The synthesis of

glycogen in the liver by the processes of glycogenesis and glyconeogenesis is counteracted by the conversion of glycogen to glucose, or glycogenolysis, and the degradation of glycogen to lactic acid, or glycolysis.

[W.Z.H.]

Glycol

One of a class of compounds having two hydroxyl (—OH) groups on separate carbon atoms of an aliphatic carbon skeleton. They undergo most of the typical chemical reactions of monohydric alcohols as well as other reactions involving both hydroxyl groups. 1,2-Glycols are produced by the hydrolysis of epoxides, by catalytic hydrogenation of aldols, by the alkaline hydrolysis of dichloroalkanes, and by reductive coupling of carbonyl compounds. The yearly production of glycols in this country exceeds 1,000,000,000 lb, with ethylene glycol accounting for the largest part of this total.

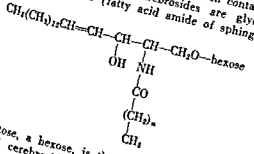
1,2-Propylene glycol, which is made by the hydration of propylene oxide, finds wide usage in antifreezes, plasticizers, tobacco humectants, food products, pharmaceuticals, cosmetics, and inks. It is an excellent solvent for many drugs, dyes, and perfumes. 1,4-Butanediol, from acetylene, formaldehyde, and hydrogen, has been used for the manufacture of butadiene (the 1,3- and 2,3-butanediols also can be used). 1,5-Pentanediol, a plasticizer intermediate, is made by the reduction of glutaraldehyde. The catalytic hydrogenation of butyraldol gives 2-ethyl-1,3-hexanediol, which is an excellent insect repellent. Thiodiethylene glycol reacts with hydrogen chloride to form mustard gas, and 3,6-dimethyl-4-octyne-3,6-diol finds use as a soporific drug. See ALCOHOL; ETHYLENE GLYCOL; POLYETHYLENE GLYCOL; POLYHYDROXY ALCOHOL.

[J.T.A.]

Bibliography: G. O. Curme and F. Johnston, *Glycols*, Am. Chem. Soc. Monograph 114, 1952.

Glycolipid

One of a class of complex lipids which contains carbohydrate residues. Cerebrosides are glycosides of ceramide (fatty acid amide of sphingosine).



Galactose, a hexose, is the major constituent of animal cerebrosides. Glucocerebrosides are also found normally and accumulate in large amounts in certain pathological conditions. A glycoside of ceramide containing both galactose and glucose has been isolated from a human carcinoma. A glucocerebroside has been found in wheat-flour lipids and a 6-sulfogalactocerebroside is also present in the brain.

Ganglioside, hematoside, and globoside are glycosides of ceramide containing several sugar residues. Ganglioside, which is found in the brain, contains glucose, galactose, galactosamine, and neuraminic acid as sugar residue. Hematoside, which is found in horse and dog erythrocytes, has similar sugar components, while globoside from human, sheep, and hog erythrocytes differs from the other two in that it contains no neuraminic acid. The structures of these compounds are not well established.

Phytoglycolipid is a recently characterized phytosphingosine-containing glycolipid from plant seeds. The sugar residues present are inositol, glucuronic acid, mannose, glucosamine, arabinose, and galactose.

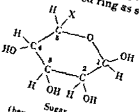
The cord factor is a toxic glycolipid (trehalose-6,6′-dimycolate) found in *Mycobacterium tuberculosis.*

Galactosyl-glycerol lipids, are mono- and digalactosyl derivatives of diglycerides found in wheat-flour lipids. See LIPID; SPHINGOLIPID.

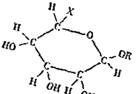
[H.E.C.; R.H.C.]

Glycoside

One of a group of derivatives of the cyclic forms of the sugars in which the hydrogen of the hemiacetal hydroxyl group has been replaced with an alkyl or aryl group, called the aglycon group. The most prevalent groups are the pentoses and hexoses, both of which exist predominantly in the form of a six-membered ring as shown:



Sugar
(hexose, $X = CH_2OH$,
pentose, $X = H$)



Glycoside
(R = aglycon)

In the hexose series, carbon atoms 1, 2, 3, 4, and 5 are all asymmetric, that is, capable of optical isomerism. Any one sugar, such as D-glucose, has a particular fixed orientation at carbon atoms 2, 3, 4, and 5, and the remaining two possible orientations at carbon 1 then represent α- and β-D-glucoses. The α form is the more dextrorotatory. In the glycosides, the group R may be aliphatic, aromatic, hydroaromatic, or heterocyclic, and it may be linked through carbon-oxygen-carbon bonds. Derived from each sugar is a pair of α- and β-glycosides that differ only in their orientations at carbon 1. The α forms of the glycosides are the more dextrorotatory, also. Glycosides of similar five- and seven-membered ring systems are also known. See MONOSACCHARIDE.

Preparation. Glycosides may be prepared by condensing the sugar with an alcohol or phenol (ROH) in the presence of a catalyst. One simple method of preparation is to dissolve the sugar in the anhydrous alcohol containing 0.5–1.0% hydro-

gen chloride and either to reflux the mixture for several hours or to allow it to stand for one or two days at room temperature. This type of procedure yields mixtures of α- and β-glycosides, but more highly specialized procedures are available for synthesis of the pure isomers. Certain enzyme systems may be used both to synthesize and hydrolyze glycosides. The glycosides are hydrolyzed by aqueous acids to form the original sugar and the alcohol or phenol, ROH. The ease of hydrolysis depends on the nature of the aglycon group, the orientation of hydroxyl groups in the sugar ring, and the size of the sugar ring. Hydrolysis of five- and seven-membered rings occurs readily and also when the hydroxyl group of carbon 2 is replaced by hydrogen (2-deoxysugars). Aromatic glycosides (R = phenyl) are also hydrolyzed by aqueous alkali, but most aliphatic glycosides ($R = CH_3$) are stable to alkali except at high temperatures (above 170°C).

Natural glycosides. Biologically the naturally occurring glycosides are one of the most important groups of the carbohydrates. Many of the colored pigments of flowers, natural dyestuffs, aromatic principles, poisons, and drugs such as the heart stimulants (cardiac glycosides), are glycosidic in nature. Glycosides are also found in animal tissues, particularly in brain tissue (cerebrosides) and urine. It has been suggested that the function of the

bilize

aglyco

ical interest in glycosides resides in the chemistry of the aglycon group. The glycosides can be classified by reference to the type of aglycon group present in the molecule.

Anthoxanthin glycosides. These are common plant pigments that consist of a sugar, usually glucose, and an anthoxanthin such as the flavones, flavanols, flavanones, isoflavones, xanthones, and anthocyanidins. Many were used as dyestuffs before the development of the synthetic coal-tar-based dyes. Certain flavone glycosides may be used to decrease the fragility and permeability of capillary walls. *See* ANTHOCYANIN; FLAVONE.

Indican. This is a naturally occurring glycoside of indoxyl and glucose. Indoxyl is oxidized by air to indigo, a dye which was originally prepared in this way, after enzymatic hydrolysis of the extracted indican.

Phenanthrene (steroid) glycosides. Two principal types of phenanthrene-related glycosides are the cardiac glycosides and the saponins. Cardiac glycosides are of medical interest because of their stimulant effect on the heart. Several of these also have been used as arrow poisons. Their most important sources are *Strophanthus* and *Digitalis* (foxglove); the latter provides therapeutically valuable drugs. Most of the cardiac glycosides have an aglycon group derived from phenanthrene, attached through one or more deoxy sugars to a normal sugar, usually glucose:

$$(\text{Glucose})_x \xrightarrow{(a)} (\text{Deoxy sugar})_y \xrightarrow{(b)} \text{Aglycon}$$

The linkage (*b*) is more readily hydrolyzed by acid than (*a*). Several synthetic glycosides of adrenal corticosteroids and other sterols have been prepared, and appear to retain the physiological activity of the aglycon with increased solubility.

The saponins are very widely distributed in plants. They are strong foaming agents and are poisonous and hemolytic when given intravenously, but usually are nontoxic when administered orally. Hence, they are used as fish poisons and the fish so killed are edible. The neutral saponins (digitalis saponins) have aglycons (sapogenins) derived from cyclopentanoperhydrophenanthrene, and the acid saponins have aglycons derived from triterpenes. In addition to the cardiac glycosides and saponins, the solanum alkaloids also contain the steroid nucleus as an aglycon. The steroid glycosides are of potential use as a source of sterols for synthesis of hormones. *See* STEROID.

Phenyl glycosides. Arbutin, salicin, and related glycosides, common constituents of leaves and barks, have substituted phenyl aglycon groups. Salicin is used as a remedy for fevers and rheumatism. Gallic acid is present in tannins as the aglycon.

Vanillin and coumarin glycosides. Many plants contain glycosidic aromatic principles that have vanillin and coumaric acid as aglycon groups. Vanillin is the aromatic principle of commercial vanilla extract, and is released from the glucoside (glucovanillin) by enzymatic hydrolysis during curing of vanilla beans.

Cyanogenetic glycosides. Amygdalin is the effective principle of oil of bitter almonds; it was one of the earliest recognized glycosides and has the structure:

$$\text{Glucose—Glucose—O—CH(CN)}C_6H_5$$

Hydroxyanthraquinone glycosides. These are found in many plant materials, and the aglycons derived from them have often been used as dyestuffs. The mordant dyestuff known as turkey red, or alizarin, is the aglycon of ruberythric acid, and it is synthesized from coal tar; formerly, it was obtained from madder, the root of *Rubia tinctoria*. *See* ANTHRAQUINONE PIGMENTS.

Streptomycin. This is an important antibiotic produced by a soil microorganism; it has the structure:

$$N\text{-Methyl-L-glucosamine} \xrightarrow{(a)} \text{Streptose} \xrightarrow{(b)} \text{Streptidine}$$

Linkages (*a*) and (*b*) are glycosidic; streptose is a substituted pentose, 5-deoxy-3-formyl-L-lyxose, and streptidine is an inositol derivative. The related streptomycin B is also a glycoside. *See* STREPTOMYCIN.

Cerebrosides. This complex group of glycosides, found in nerve tissue, normally consists of a hexose (galactose or glucose), a nitrogenous base, and a fatty acid. The related gangliosides contain several sugar units in each molecule, and often include amino sugars. *See* CARBOHYDRATE; GLUCOSE; LIPID. *See also* GLUCOSIDE. [C.N.R.; R.L.WH.]

Gnathiidea

A suborder of the Isopoda. This group of crustaceans was confused with the Flabellifera until it was made a suborder by H. J. Hansen in 1916. These animals are characterized as having a much reduced second thoracomere which is incorporated with the cephalothorax. The antennules are short and each has a flagellum with 4–5 joints; the terminal three bear single sensory palps. The antennae are also short, with flagella having 5–8 joints, but generally there are seven. Thoracomeres 3–7 have a normal appearance and bear pereiopods, while the eighth thoracomere is vestigial and lacks appendages. The pleon is straight. It has 5 pleomeres and a triangular pleotelson. Both pleopods and uropods are foliaceous.

The male is more or less flattened dorsoventrally, with the large, broad cephalothorax often quadrangular in shape. These animals lack maxillules, the maxillae are vestigial, and maxillipeds are present. The second thoracopod, or pylopod, is operculate, while an appendix masculina occurs rarely.

The female has an ovoid appearance with a very small cephalothorax. The mouthparts are greatly reduced or absent, and the thoracopod is not operculiform. Development occurs in a brood pouch in which an ovary becomes a uterus.

The larva has been described under the name *Praniza*. It resembles the female but is more fusiform. The mouthparts are of the piercing and sucking type. The second thoracopod is modified as a gnathopod with a strong terminal hook. The larvae are found as parasites on fish. *Praniza milloti* Monod, a parasite of the coelocanth, has a pediform gnathopod which is a primitive condition.

The Gnathiidea undergo a true metamorphosis. Their extreme polymorphism led to the separation of the males (form *Anceus*) from the females and larvae (form *Praniza*) as separate genera. The group comprises marine forms exclusively. They are found in all latitudes and at all depths. As a group, they occupy a place that is distinctly apart from the other Isopoda. See FLABELLIFERA; ISOPODA.

Bibliography: T. Monod, *Les Gn[...] monographique (morphologie, [...] [...]tique)*, Mém. soc. sci. nat. [...]

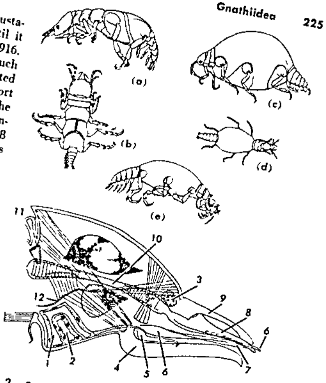


Fig. 2. *Paragnathia formica* (Hesse). (a) Male, lateral view. (b) Male, dorsal view. (c) Female, lateral view. (d) Larva (*Praniza*), dorsal view. (e) Larva (*Praniza*), lateral view (after T. Monod, *Mém. soc. sci. nat. Maroc 13, 1926*). (f) Cephalon of the larva (*Praniza*). 1, duct of maxillary gland, 2, terminal sac; 3, paraantennary gland, 4, maxilliped; 5, maxilla; 6, maxillule; 7, paragnathus; 8, mandible; 9, labrum; 10, esophagus, 11, tergite of the gnathopodial somite; 12, tentorium.

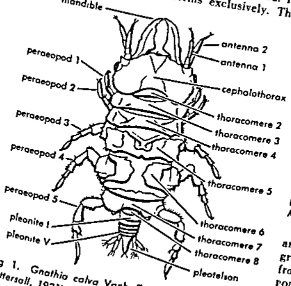


Fig 1. *Gnathia calva* Vanhoffen, male. (After W. [illegible] Tattersall, 1921)

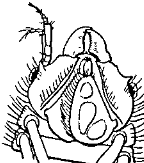


Fig. 3 *Gnathia antarctica* (Studer), male, ventral surface of anterior showing the operculate pylopodes (appendages of second thoracic somite). (After T. Monod, *Mém. soc. sci. nat. Maroc 13, 1926*)

Gnathostomata

A group of the subphylum Vertebrata which possess jaws and usually have paired appendages in contrast to the Agnatha which lack jaws and paired appendages. The Gnathostomata are divided into two main subdivisions, the superclasses Pisces and Tetrapoda. See AGNATHA; PISCES (ZOOLOGY); TETRAPODA. [C.B.C.]

Gneiss

An old term in mining and geology that was first applied by the miners of the Erzgebirge, Austria, to designate the country rock adjacent to the ore veins. Gneiss still has a broad definition and in its widest sense it is used as a structural rather than compositional term. Thus it applies to a great variety of rocks with a banded or coarsely foliated structure. Precambrian gneiss was once regarded as the material of the first primitive crust of the earth. A modern version of this is to include in the definition of gneiss (primary gneiss) that it is of plutonic igneous origin with banding caused by flow movements in a crystallizing heterogeneous magma.

Typical gneisses are highly metamorphic, coarse-grained, irregularly banded or foliated rocks, composed mainly of quartz and feldspar with some biotite (= granitic composition). As the grain becomes finer and the amount of mica increases, gneisses pass into pelitic schists. With fine grain and decrease of mica, or garnet substituting for mica, they become less foliated and pass into granulite. See GRANULITE; SCHIST.

of folded mountains a pore fluid, or ichor, is generated by differential melting (palingenesis or anatexis), and the solid residues and the contiguous sediments are stewed in this magmatic-anatectic liquid which reacts with the sediments and metasomatically transforms them into migmatic gneisses. Lit par lit gneisses and augengneisses belong to this category. See METASOMATISM; MIGMATITE.

Paragneiss shows a sedimentary parentage. In the deeper parts of the orogenic belts any sedimentary complex will recrystallize under rather high temperature and pressure (the conditions of the amphibolite or the granulite facies). Aided by emanations or ichors of granitic composition, various types of paragneisses may form. If derived from argillites, the gneisses may be rich in muscovite, biotite, garnet (almandine), cordierite, kyanite, or sillimanite. If calcareous material is present, the gneisses may carry epidote, augite, garnet (grossularite), and wollastonite. Many banded gneisses have a sedimentary parentage.

Orthogn

terms like gabbro gneiss have been used. Ordinarily the term gneiss refers to a granitic composition; anything else should be specially noted. Typical gneisses (of granitic composition) make up large areas in Precambrian regions and in the deeper parts of orogenic belts. See OROGENY. [T.F.W.B.]

Gnetales

An order of the plant class Gymnospermae represented by three genera: *Ephedra*, *Welwitschia*, and *Gnetum*. *Ephedra* has some 30 species, mostly straggling, xerophytic shrubs in dry, rocky, or desert regions. The leaves are opposite and small and rudimentary (see EPHEDRA). *Welwitschia*, having only one species, is a bizarre seed plant. The tuberous stem, about 18 in. long, is often nearly covered with soil. It has a long tap root and supports two thick, leathery, strapshaped, persistent leaves. *Welwitschia* occurs only in the western part of South Africa where the average annual rainfall, supplemented by heavy dews, is only ⅓ in. *Gnetum* has 34 species, all inhabitants of the tropics. These are prevailingly lianas (climbing plants), with only a few shrubs and trees. The large, oval, entire, netted-veined leaves, resembling dicotyledon leaves, are opposite and decussate (arranged in pairs each at right angles to the next pair above or below). There are 19 Malaysian, 2 African, 7 South American, and 1 West Indian species. There are none in North America, Europe, or Australia.

The origin and relationships of the Gnetales are uncertain. Because all have naked ovules, they are unquestionably gymnosperms. However, the presence of true vessels in the secondary wood and compound strobili (cones), both male and female, separate them from all other gymnosperms. These two features, combined with opposite leaves, dicotyledon embryos, no resin canals, and the long micropylar tube formed by prolongation of the inner integument, bind the three genera together in one order. The presence of both gymnosperm and angiosperm characteristics suggests that the Gnetales have evolved from the Coniferales. Because of their angiosperm features, the Gnetales have been regarded as possible ancestors of the angiosperms. This seems unlikely since no Gnetales (fossils) occur below the Tertiary, whereas angiosperm fossils are abundant in the earlier Cretaceous (see GEOLOGY). See PLANT KINGDOM. [P.D.S.]

Bibliography: See EMBRYOPHYTA.

Gnu

Any of three, and possibly more, species of African antelopes, also known as wildebeests, of the genus *Connochaetes*, family Bovidae. Gnus have been described as looking like a horse with horns. The hindquarters, legs, and long flowing tail and mane make the gnu resemble a horse, but the head, bearing curved horns and a broad muzzle, is cowlike.

Gnus are frequently found in herds in close association with zebras. Whole herds are apt to erupt in a display of coltish antics at the first indication

The [illegible], Connochaetes taurinus; length to 7 ft. (From P. Martin Duncan, ed., Cassell's Natural History, Cassell & Company, Ltd., London)

of danger or disturbance. Gnus are a favorite food for humans and wild predators alike, and are rapidly being reduced in numbers. *See* ARTIODACTYLA; ZEBRA.

[J.D.B.]

Goat

Any of several ruminants of the family Bovidae. Domestic goats are of Old World origin and belong to the genus *Capra*. The angora goat, *C. angorensis*, is raised throughout the world in a number of varieties, primarily for its wool. The milk goat, *C. hircus*, is also widely reared as a domestic animal. The mountain goat, *Oreamnos americanus*, is a prized sport trophy which inhabits the more inaccessible portions of the western American mountains from Idaho to Alaska. *See* ARTIODACTYLA.

[J.D.B.]

GOAT PRODUCTION

Range goats. In the United States, range goats are concentrated in the Southwest, particularly in the Edwards Plateau of Texas. Of the 2,800,000 goats in the nation, 81% are Angora (Fig. 1). Of these, 95% are in Texas. Common or Spanish goats (no particular breeding) as well as Angoras are utilized for clearing brushy land (Fig. 2). Under such circumstances goats do not replace cattle; they are added to the carrying capacity of the range to browse oak sprouts and other woody nuisance plants (Fig. 3).

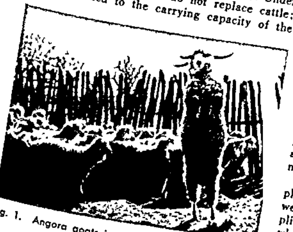


Fig. 1. Angora goats in corral

Angora goats were imported to the United States from Turkey as early as 1849. This breed produces lustrous white mohair (*see* MOHAIR). Mature Angora bucks have a range in weight of 125–225 lb; does, from 70–110 lb; and wethers (castrated rams), 75–125 lb. These weights vary in relation to range conditions.

Spanish goats are larger than Angoras, but their short, brownish-black coats have no market value and they are not used as milk producers.

Milking Goats. Four breeds of milking goats predominate: the Nubian, Toggenburg, French Alpine, and the Saanens. These are concentrated around heavily populated areas on small farms. Goat milk is a specialty item of particular value for infants and invalids with stomach disorders because its butterfat is easily digested.

…ngora goats feeding on brush.

Marketing. Surplus goats are marketed through the usual livestock channels, with major markets in San Antonio and Fort Worth, Texas. The meat is called chevon or cabrito. The largest quantities are consumed in the Southwest, being especially savored as barbecue.

Breeding of Angora goats. Efficient herd management at breeding time is profitable. Preparation for breeding includes conditioning the bucks and flushing (extra feeding) the does, shearing impeding locks of mohair, and culling inferior does.

Newly purchased bucks should be given time to adjust to any change in environment. Health and vigor should be insured by feed supplements, such as whole yellow corn or 41% protein cottonseed pellets (*see* CORN; COTTON; PROTEIN). Oats and alfalfa are also used with or without cottonseed meal (*see* ALFALFA; OATS).

Does breed more efficiently if they are flushed by placing them in a fresh well-vegetated pasture 2–3 weeks prior to breeding. The same result is accomplished by feeding a supplement, such as yellow whole corn and 41% cottonseed pellets. These may be fed on the ground. In troughs, 0.25–0.50 lb cottonseed meal or ground alfalfa is excell

Fig. 3. Results of "goating" cedar

Fig 4. Kidding method, staking in the pen. *(Sheep and Goat Raiser Magazine, Texas)*

Well-grown does and bucks are mature for breeding at 17 months. Bucks are bred at the rate of about 1 buck to each 30–35 does. The breeding season is chiefly in October and November; however, does come into heat during late summer, fall, and early winter. Bucks are fertile throughout the year. Does remain in estrus 1 or 2 days; the estrous cycle varies from 17 to 21 days; gestation ranges from 147 to 155 days. Bucks should be with does for 40 days, permitting those does not conceiving during the first heat to be bred again during the second period. Milk goats follow the same breeding habits as Angoras; however, common goats breed throughout the year and can have two crops of young each year.

Kidding. Most Angora kids are born in March on fresh pasture. Ranchmen practice three methods of kidding: loose range, loose pen, and staking in the pen (Fig. 4).

Loose-range kidding is hazardous as the kids are dropped haphazardly over the range without the individual attention to the kid and mother that is advisable.

In the loose-pen system, does are turned on the range during the day. As kids are born, they are brought into the pen, remaining there for a month. The mothers graze alone on the range during the day and return to the pens at night to nurse their kids.

Some ranchers practice staking kids in the pen. In this system three kinds of pens are used: a feed pen (holding 250 does); a stake pen (25 does and kids); and a loose or free pen (250 does and kids). Frequently a shed common to all pens provides shelter from inclement weather.

Selection. Angora goats have not been studied extensively by research workers. There are no her-

Scoring card for Angora goats

Characteristics	Scoring points, %
Body (40 points)	
Size and weight for age Min weight yearling buck 80 lb; yearling doe, 60 lb	11
Constitution and vigor Width and depth of chest; fullness of heart-girth, and spring of ribs	11
Conformation Width and depth of body; straightness of back, width of loin, strength of back	8
Amount of bone Indicated by size of bone below knee and hock (should be clean and in proportion to size of animal); strength of feet and legs	5
Angora breed type Indicated by head, horns, ears, color markings (freckles not objectionable); horns should be wide-set on bucks and should spiral outward	5
Fleece (60 points)	
Freedom from kemp Kemp is large, white, chalky hair	12
Uniformity and completeness of covering Includes uniformity of fineness, length, type of lock and covering [illegible]	11
[illegible]	9
Density Indicated by number of fibers per unit of area; visually, by amount of skin exposed when fleece is parted	8
Length of fleece Equivalent to 1 in. or more growth per month	8
[illegible]	8
[illegible] ni- [illegible] let, or web)	4
Total	100

itability estimates except for cryptorchidism (defect of the testes). Breed improvement has been attained by breeders using empirical methods.

Judging. The Texas Angora Goat Raisers Association has published a guide for judging Angora goats. (See scoring card.) Body size in Angora goats is related to strength and vitality. Size, however, is frequently subordinated to fleece quality (fiber diameter). Overemphasis on selection for fleece quality tends to reduce size and vigor.

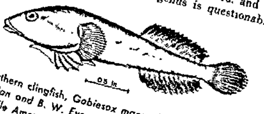
A good Angora goat produces 1 in. of mohair each month. As they are sheared twice annually, a clip should be 6 in. long. A ready market exists for the fine mohair but coarse mohair is sometimes difficult to sell. Uniformity of fiber diameter within fleeces is important. The coarsest fiber grows beneath the chin of Angora goats. This is a good location to check uniformity of fiber diameter. As a general rule, mohair tends to be coarser on the shoulder and finer on the breech (hinder part).

For registry, disqualifying characteristics include deformed mouths, broken-down pasterns (foot bones), deformed feet, crooked legs with large hocks, wool-like fleeces, abnormalities of testicles, close-set and distorted horns, and any off-color hair. See SHEEP; WOOL. [T.D.W.]

Bibliography: See AGRICULTURAL SCIENCE (ANIMAL).

Gobiesociformes

An order of bony fishes, also known as the Xenopterygii, equipped with a thoracic sucking disk which serves for attachment to the bottom. This papillose disk involves the four rays of each pelvic fin, pelvic and pectoral girdles, and dermal flaps. There are single dorsal and anal fins that lack fin spines The body is scaleless, and there are no ribs and no swimbladder. The posttemporal is not forked. The order consists of a single family that is classified into 8 subfamilies, 33 genera, and 93 Recent species. A Miocene genus is questionably assigned to the family. The Gobiesocidae are interpreted as perciform derivatives. Clingfishes inhabit the intertidal zone of tropical to temperate shores of all continents, and a few occur in swift coastal streams of the American tropics. See ACTINOPTERYGII. [R.M.B.]



Northern clingfish, *Gobiesox maeandricus.* (After D. S. Jordan and B. W. Evermann, The Fishes of North and Middle America, U.S. Natl. Museum Bull. 47, 1900)

Bibliography: J. C. Briggs, *A Monograph of the Clingfishes (Order Xenopterygii)*, Stanford Ichthyological Bull., vol. 6, 1955.

Goethite

A mineral having composition FeO(OH) and comprising most of the material known as brown iron ore. Goethite crystallizes in the orthorhombic system. Crystals are rare and the mineral is usually in radiating fibrous aggregates with a reniform surface. It may also be stalactic or massive. There is one perfect cleavage. The hardness is 5–5.5 (Mohs scale); the specific gravity is 4.37, or lower because of impurities. The luster is adamantine to dull and the color light to dark brown. It is characterized by a yellowish-brown streak.

Limonite commonly appears to be amorphous but is similar in its physical and chemical properties and occurrence to crystalline goethite. Until the study of these minerals by x-rays in the 1920s, goethite was considered to be a relatively rare mineral and limonite common. X-ray analysis showed most of the presumed limonite to be crystalline and therefore goethite.

Goethite is a common mineral formed by the oxidation of iron-bearing minerals. It is the major constituent of the gossan at the surface of metalliferous veins, forms as the residual mantle from the weathering of serpentine, and is deposited as bog-iron in marshes and stagnant pools. In some localities it is found in large masses and constitutes a valuable iron ore, for example, the minette ores of Alsace-Lorraine. See IRON (EXTRACTION FROM ORE); LIMONITE; ORE AND MINERAL DEPOSITS. [C.S.HU.]

Goiter

An enlargement of all or part of the thyroid gland, due to various causes, which may be accompanied by a hormonal dysfunction.

Colloid, endemic, or simple goiter results from an iodine deficiency. Iodine is a necessary precursor of the thyroid hormone, thyroxine. In this group of diseases a decreased thyroxine in the blood from lack of iodine causes the pituitary to increase its thyrotrophic hormone output. The latter, in turn, produces hyperplasia and increased function so that the iodine is used more fully in the attempt to attain euthyroidism, or a balanced output. Endemic goiter is peculiar to many geographic areas in the world where iodine is lacking in food, soil, or water; most of a local population may show evidence of this or related thyroid dysfunction.

Diffuse primary hyperplasia, or exopthalmic goiter, results from an over-all glandular enlargement accompanied almost always by excess thyroxine secretion. The exact cause of this disorder, also called Grave's disease, is unknown but may be due to overstimulation by a deranged pituitary.

Nodular goiter, or multiple colloid goiter, is usually the result of repeated cyclic changes in a part of the gland subjected to other abnormal processes such as the two preceding conditions. Thyroid output may be decreased but more often is normal or increased.

Other enlargements are not true goiters, but externally may be indisting

These result from various infections or from tumor formations of either benign or malignant types. Occasionally, certain drugs may cause thyroid enlargement.

Goiter does not necessarily mean thyroid dysfunction, since enlargement may be seen in hypothyroid, euthyroid, and hyperthyroid states. Conversely, dysfunction may occur in a gland that shows no visible change. See HORMONE; THYROID GLAND. [E.G.ST.]

Gold

Chemical element number 79, gold, Au, is a deep yellow, soft, and very dense metal. Gold is classed as a heavy metal and as a noble metal; commercially, it is the most familiar of the precious metals. Copper, silver, and gold comprise group Ib of the periodic table of elements.

There is only one stable isotope of gold, that of mass number 197. Some 20 radioactive *isotopes* have mass numbers from 187 to 203.

Throughout history, gold has been known and prized. In 353 B.C., Xenophon wrote that the earliest records of Grecian gold mines were lost in antiquity. Gold has been found in Egyptian tombs from the Stone Age.

Uses of the metal. Copper, silver, and gold are sometimes called the coinage metals because of their well-known usefulness in coins. Gold *coins*, as well as most other objects containing gold, are actually gold alloys, because the metal itself is too soft to be very useful. Some architectural *gold leaf* is nearly pure gold, as are some items of chemical apparatus. Gold sols have been used in making colored glass. A suspension of gold in oil, called liquid gold, is *sometimes used* for decorating china. Proof gold, which is 99.99+% pure, is used for certain chemical and physical standardizations. See GOLD ALLOYS.

Occurrence. Gold occurs widely throughout the world, but usually *very sparsely*, so that it is quite a rare element. Sea water contains some gold (but less than was once believed), perhaps 0.0005 parts of gold per million parts of water. Somewhat higher concentrations *accumulate on* plankton or on the ocean bottom. Native, or metallic, gold and various telluride minerals are the only forms of gold found on land. Native gold *may occur in* veins among rocks and ores of other metals, or it may be scattered in sands and gravels (alluvial gold). The gold nuggets found in the western United States in the nineteenth century were typical alluvial deposits. One alluvial nugget found in Australia weighed about 600 lb. Most native gold contains silver in amounts ranging from 1 to 50%. Telluride minerals are less common than native gold; among them are calaverite ($AuTe_2$), sylvanite ($AuAgTe_4$), and petzite [$(Au,Ag)_2Te$]. See GOLD METALLURGY.

Gold metal. The density of gold is 19.3 times that of water at 20°C, so that 1 ft^3 of gold would weigh about 1200 lb. Gold melts at 1063°C and boils at about 2700°C. It is somewhat volatile well below its boiling point; mixtures of platinum and gold can be separated by keeping them melted for a long time to drive off the gold. Gold is the most malleable and ductile of all metals. It can easily be made into translucent sheets 0.00001 mm thick or drawn into wire weighing only 0.5 mg/meter. The quality of gold is expressed on the fineness scale as parts of pure gold per thousand parts of total metal, or on the carat scale as parts of pure gold per 24 parts of total metal. Gold and silver form true solutions (alloys) in any proportions. Gold readily dissolves in mercury to form amalgams.

Chemical properties. Gold is one of the least active metals chemically. It does not tarnish or burn in air. It is inert to strong alkaline solutions and to all pure acids except selenic acid. In order to dissolve gold chemically, it is best to use a combination of oxidizing power and complexing ability such as is found in a mixture of nitric and hydrochloric acids, called aqua regia because it can dissolve the royal metal. A similar effect occurs when gold dissolves in solutions of cyanide ion (a complexing agent) containing oxygen (the oxidizing agent), forming $[Au(CN)_2]^-$ complex ions. Gold also reacts with bromine at room temperature and with fluorine, chlorine, iodine, and tellurium at higher temperatures. One of the interesting characteristics of gold is its ability to exist in a sol, or colloid. Depending upon the particle size of the gold, aqueous gold sols may be red, *blue, or purple*. Tannin, formaldehyde, and phenylhydrazine are some of the common reducing agents added to solutions of gold compounds to generate the sols. The beautiful sol known as Purple of Cassius is generated by addition of tin(II) chloride. This sol may possibly contain complexes of gold instead of or in addition to colloidal gold.

Gold compounds. Gold may be either unipositive or tripositive in its compounds. So strong is the tendency for gold to form complexes that all the compounds of the 3+ oxidation state are complex. The compounds of the 1+ oxidation state are not very stable and tend to be oxidized to the 3+ state or reduced to metallic gold. All compounds of either oxidation state are easy to reduce to the metal. This situation follows the usual rule that compounds of inactive (noble) metals readily revert to the metals, whereas the reduction of compounds of active metals is difficult.

Compounds of gold

Name	Formula	Properties, uses, remarks
Tetrachloroauric acid 4-hydrate; *chloroauric* acid; gold chloride	$HAuCl_4 \cdot 4H_2O$	Soluble complex compound; solutions are a common commercial *form of combined gold*; used to prepare other gold compounds, gold sols
Sodium tetrachloroaurate 2-hydrate	$NaAuCl_4 \cdot 2H_2O$	Made from $HAuCl_4$ solutions, soluble compound; used in gold plating, photographic *tinting, and some* medicines
Gold(I) cyanide	AuCN	Does not exist as single AuCN molecules, but as long chains of alternate gold atoms and CN groups
Cesium hexachlorodiaurate	$Cs_2Au_2Cl_6$	Contains gold in both 1+ and 3+ oxidation states; originally thought to be a compound of gold in 2+ state
Gold(III) chloride	$AuCl_3$	Red soluble compound, made by reaction of gold and chlorine, or by reaction of $HAuCl_4$ with chlorine; forms $HAuCl_4$ in hydrochloric acid solution; with ammonia, forms explosive fulminating gold
Gold(III) hydroxide	$Au(OH)_3$	Yellow-brown insoluble compound, formed by addition of hydroxide ion to $AuCl_3$ or $HAuCl_4$ solution; may be a hydrous form of the oxide Au_2O_3; dissolves in most acids; easily reduced to metallic gold; is amphoteric, dissolving in excess hydroxide ion to form complex hydroxoaurate ions

In its complex compounds gold forms bonds most readily and stably with halogens and sulfur, less stably with oxygen and phosphorus, and only weakly with nitrogen. Bonds between gold and carbon are fairly stable, as in the cyanide complexes and a variety of organogold compounds. The complexes of 1+ gold are usually linear, having two groups attached to the metal atom in diametrically opposite positions. Complexes of 3+ gold are square planar, having four groups attached to the metal.

Analytical methods. Qualitative detection of gold in solution is usually accomplished by addition of reducing agents to liberate metallic gold which is identifiable either by the yellow color of the compact metal or by a characteristic color of a gold sol. Two color reactions of gold that are particularly *sensitive in solution are those with* dithizone and benzidine. Gold may be separated from the platinum group metals by precipitation with hydroquinone in dilute hydrochloric acid. Quantitative analysis for gold usually involves reduction to the metal and weighing. One volumetric procedure uses the reaction between tetrachloroaurate ion and iodide ion. The gold is reduced from the 3+ state to the 1+ state, and the iodine released is titrated with standard sodium thiosulfate solution. *There is a special fire assay method for the evaluation* of gold ores. *See* ASSAYING. [W.E.C.]

Bibliography: J. C. Bailar, Jr. (ed.), *The Chemistry of the Coordination Compounds*, 1956; N. A. Lange and G. M. Forker (eds.), *Handbook of Chemistry*, 9th ed., 1956; N. V. Sidgwick, *The Chemical Elements and Their Compounds*, vol. 1, 1950; M. C. Sneed, J. L. Maynard, and R. C. Brasted (eds.), *Comprehensive Inorganic Chemistry*, vol. 2, 1953.

Gold alloys

Combinations of gold and other metals. Gold is not oxidized in air, and for thousands of years was the only metal available that remained bright indefinitely. This, coupled with its pleasing yellow color, made it cherished for personal adornment and other decorative uses and valuable as a medium of exchange. The high specific gravity of gold (19.3) also gave some insurance against dilution with cheaper metals of lower density. Pure gold is very soft. The addition of copper hardens the gold, and ultimately, gold-copper alloys became standard for coinage. Gold coins in the United States contained 10% copper, the balance gold. In recent years, most of the world's gold has been confined in vaults here and there, and is used only for international government transactions.

Although gold is widely distributed, most of the ores are very lean, so that its production is costly and as a result the price of gold has always been high in relation to that of other metals. At present, *it is priced at $35/troy oz in the United States, but* it cannot be purchased freely except for industrial or jewelry uses.

About 40,000,000 oz of gold per year, worth nearly $1,400,000,000 are now being produced in the world. Nearly half of this comes from South Africa. The largest producer in the United States has long been *Homestake in South Dakota. Only* about 10% of the world production of gold is used in industry; the United States uses about 1,500,000 oz, one-half of which goes into jewelry.

Jewelry. Pure gold is very weak, having a ten-

is designated in carats (abbreviated Kt); pure gold is 24 Kt, 18 Kt is 18/24 or 75% pure gold, and 14 Kt is 14/24 or 58.3% pure gold.

Very early, it was found that reasonably pu
gold could be hammered into extremely thi
ut 0.000005 in. thick, and this was

architectural and other decorations in pre-Biblical times. Such foil or leaf is still made and used in about the same way.

Because of the relatively high cost of gold it is often used in laminated form, a thin layer of carat gold being welded to a supporting metal, such as nickel or brass, which may be rolled or drawn into complex forms without rupturing the gold alloy surface. Material of this nature is called gold-filled or rolled-gold plate, and the ratio of the weight of the gold alloy to the total weight of the composite material is expressed as a fraction, along with the carat of the gold alloy; for example, 1/10 12 Kt gold means that the article contains 5% of pure gold by weight. An exception to this method of designating quality occurs in watch cases, for [illegible]

[illegible] soluble in all single acids, although it can be dissolved in a mixture of nitric and hydrochloric acids (aqua regia) or in a solution of alkaline cyanides in the presence of air or an electric current. This solubility in cyanide is widely employed in extraction of finely divided gold from ores. This method will recover gold that has escaped amalgamation, so that many waste heaps from early mining operations have been reworked to extract additional amounts of gold. When an electric current is passed through a cyanide solution, gold can be deposited on other metals as an electroplate, and this is frequently employed for protecting and decorating base-metal jewelry.

The industrial uses of gold depend primarily upon the corrosion-resistance of gold and the high-gold alloys, and secondarily upon the strength that can be secured by alloying alone or alloying and heat treating. Many of the alloys used in dentistry contain gold, silver, and copper, often with small amounts of platinum and palladium, and these can be heat-treated to develop strengths above 150,000 psi, with good spring properties. Of course, they also possess a high level of corrosion- and tarnish-resistance. Alloys of this same nature find many electrical uses as contacts, particularly where rubbing is involved. High-gold alloys are also used for make-and-break electrical contacts, at currents less than ½ ampere, such as those employed in many instruments and certain telephone equipment. In the latter application, gold alloys have largely been replaced by palladium which is more economical and more resistant to electrical erosion. Gold electroplate, often very thin, is employed on high-frequency conductors such as those in radar equipment because of the high electrical conductivity of gold. Thin coatings applied to metallic or glass surfaces may be employed for reflectors, particularly for infrared radiation.

Gold, which melts at a temperature intermediate between the melting points of silver and copper, finds some use, with its alloys, as a corrosion-resistant brazing material in chemical equipment and also in certain electronic vacuum devices. A gold coating is also sometimes applied to the grids of vacuum tubes to minimize electron emission from the grid.

Because gold does not oxidize when heated in air, appropriate compounds can be decomposed by heating, with the liberation of gold. Compounds of this type are used in the decoration of china and also for the production of printed electrical circuits on ceramics. These materials, known as liquid bright golds, are applied in the form of varnish, which is dried and then heated to redness, leaving a thin film of gold firmly attached to the underlying ceramic. This coating may be as thin as 0.00004 in., but may be made thicker. It is even possible to utilize this technique for applying thin gold coatings to some of the more stable plastics, which are being used for special printed circuits. Certain gold alloys with platinum and palladium can also be applied in this manner to produce stable electrical resistors. *See* GEM MOUNTING; GOLD; GOLD METALLURGY. [E.M.WI.]

Gold metallurgy

Extraction of gold from ores, refining, and preparation for use. Treatment of gold ores yields about 30,000,000 troy oz of gold annually (exclusive of the Soviet Union, for which official figures are not available but whose annual production is estimated at about 10,000,000 oz). Some of the principal gold-producing countries (with 1957 production in ounces $\times 10^6$) are South Africa, 17.0; Canada, 4.4; United States, 1.8; and Australia, 1.1. Nearly every other country produces some gold. Only about 20% of the gold produced is consumed industrially; most of the remainder is added to monetary gold reserves. Since 1934, the price of gold in the United States has been fixed by government decree at $35/oz (fine).

Gold has been known since prehistoric times and is possibly the first metal used by man. It is widely distributed in nature, usually in metallic form, but occasionally as telluride compounds such as the mineral calaverite ($AuTe_2$). In placer deposits the gold particles may be relatively free and are often sufficiently large to be concentrated by simple physical methods based on gravity separation (for example, panning, rocking, and hydraulic mining). However, in underground ores, which are now the chief source of gold, the metal (almost invariably alloyed with some silver) is usually very finely dispersed in quartz veins or lodes, not infrequently associated with sulfide minerals such as pyrite, arsenopyrite, chalcopyrite, and galena. Such ores must be finely ground and subjected to a suitable chemical treatment for extraction of the gold.

Cyanidation. This is by far the most important process employed to extract gold from its ores. Developed in Glasgow, Scotland, in 1887 by J. S. MacArthur, R. W. Forrest, and W. Forrest, and widely used since then in most gold-producing countries, cyanidation employs a dilute solution of sodium cyanide to extract gold from the ore (which in some cases must first be subjected to preliminary

metallurgical treatment such as flotation, roasting, or both). The leaching reaction is

$$4Au + 8CN^- + O_2 + 2H_2O \rightarrow 4Au(CN)_2^- + 4OH^-$$

Continuous agitation and aeration of the pulp for a period of up to 50 hr is required for complete extraction. Lime is frequently added to control the pH and reduce the harmful effects of certain minerals, notably sulfides of copper, iron, antimony, and arsenic (cyanicides) which are commonly present and which consume cyanide and oxygen, thus causing poor gold recoveries.

After extraction of the gold, the solution is separated from the ore tailings by filtration, passed through a vacuum chamber and thus deaerated (Crowe process), and the gold is then precipitated by the addition of zinc dust, according to the reaction

$$2Au(CN)_2^- + Zn \rightarrow 2Au + Zn(CN)_4^{--}$$

The precipitate, containing gold as well as some silver and unreacted zinc, is then smelted to eliminate the zinc, and the resulting gold-silver alloy (usually containing 70–90% gold) is further refined by parting with nitric acid or hot concentrated sulfuric acid (which dissolves silver but not gold), by chlorination of the molten metal (converting silver to the chloride which is removed as a fume or scum), or by electrolysis.

Other methods of extraction. Prior to the development of cyanidation, the most important method of extracting gold from ores was amalgamation. In this process, which is still practiced on a restricted scale for ores containing coarse gold, a slurry of finely divided ore flows over a mercury-coated surface, and the gold particles form an alloy (amalgam) with the mercury. The amalgam is collected, and the gold is recovered by distillation of the mercury.

Another process formerly used to recover gold from ores, chlorination, is now obsolete.

Small quantities of gold frequently occur in ores of copper, lead, zinc, nickel, and other base metals, and are recovered as a by-product of the smelting of these ores or electrorefining of the metals. See AMALGAM; GOLD; GOLD ALLOYS; SILVER METALLURGY. [J.H.]

Bibliography: J. L. Bray, *Non-Ferrous Production Metallurgy*, 2d ed., 1947; J. V. N Dorr and F. L. Bosqui, *Cyanidation and Concentration of Gold and Silver Ores*, 2d ed., 1950; D. M. Liddell (ed.), *Handbook of Nonferrous Metallurgy*, vol. 2, 2d ed., 1945.

Goldfish

The most common of aquarium pets, *Crassius auratus*, a member of the family Cyprinidae, closely related to the carp. In nature the goldfish is brown, the golden color being a variation that has been preserved by selection. This fish, native to Asia, has been cultured there for centuries. Black, white, and mottled varieties are bred, as well as others with specialized eyes, heads, or fins. Goldfish are established in many lakes and rivers in the United States, where they commonly attain a weight of 3–4 lb. They tend to revert to the wild color and are often mistaken for carp which they resemble in habits as well as appearance. They hybridize freely with carp. Goldfish, however, do not have the barbels characteristic of carp. See CARP; CYPRINIFORMES. [J.D.B.]

The goldfish, *Crassius auratus*; length to 18 in. (From *E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

Golgi component

A minute stack of hollow laminae in the cell. They break up at their edges into a network of vesicles and channels. The form and size of the Golgi components vary; however, the stack is usually curved. The Golgi component, also known as the

Fig. 1. Semidiagrammatic drawing of portions of the Golgi zones in parts of two epithelial cells of the epididymis of the mouse. The Golgi component consists of groups of smooth surfaced flattened sacs (S) which dilate into large vacuoles (GV) Clusters of small vesicles (V) constitute the third element The cisternae (E) with attached small granules marginating the Golgi component represent the basophilic component of the cytoplasm (ergastoplasm), the membranes of which a part of the endoplasmic reticulum.

Fig. 2. Drawing of a portion of a mouse tumor cell (sarcoma 37) showing a simplified form of the Golgi component consisting of flattened sacs (S) and clusters of small vesicles (V). Occasional cisternae of ergastoplasm are also present (E) and a portion of the cell nucleus (N) appears at the right

Golgi apparatus, complex, or material was first described by C. Golgi in Purkinje cells of the cerebellum of the barn owl in 1898. Using a silver impregnation method he described a system of branching and anastomosing strands in the cytoplasm of these cells and termed it the apparto reticolare interno.

During the years since its discovery, the Golgi component has continued to command the interest of cytologists. This interest is derived partly from the enigma concerning the role of the component in cell function, and partly from the desire to know whether or not its reticular form, as seen in specimens impregnated with heavy metals, is a reasonable representation of its form in living cells. Although the negative and positive image of the Golgi component has been described repeatedly in living cells, competent observers have on occasion questioned its reality, some being convinced that the impregnated reticulum seen in fixed specimens was a gross artifact of no basic significance. Studies with the electron microscope have demonstrated that the Golgi component in actively functioning cells consists of a system of large vacuoles marginated by and probably derived from groups of flattened sacs, the latter appearing on cursory examination as layers of double membranes. Clusters of small vesicles, 40–60 mμ in diameter and formed by budding from the flattened sacs, make up the third element of this system (Fig. 1).

In anaplastic (nonfunctioning) tumor cells, the Golgi component may consist merely of a cluster of vesicles associated with an occasional flattened sac (Fig. 2). Further evidence from electron microscopy indicates that in some cases, but not in others, secretory material is formed directly within the elements of the Golgi component. R. H. Bowen's evidence for a direct relationship between the Golgi component and the formation of the acrosome of developing sperm has been confirmed, and the homology between the Golgi component of verte …). brate g … e- Study … n duction … with all three elements of the system, but it seems probable that its reduction between the flattened sacs produces the Golgi reticulum of classical cytology. High specific concentrations of phospholipid and acid phosphatase have been de-

Fig. 3. Drawing of a dictyosome of a spermatid of the earthworm showing a basic morphologic homology with the vertebrate Golgi component. In this case the flattened sacs (GS) are closely packed in a lamellar pattern. Clusters of small vesicles (V) are also present. A portion of the cell nucleus (N) is at the left.

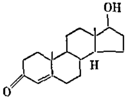
termined for the Golgi component isolated by combining tissue homogenization, and ultracentrifugation with electron microscopy. See CELL (BIOLOGICAL); CELL STRUCTURES. [A.J.D.]

Bibliography: R. H. Bowen, Cytology of glandular secretion, *Quart. Rev. Biol.*, 4(4):484–519, 1929; A. J. Dalton and M. D. Felix, A comparative study of the Golgi complex, *J. Biophys. Biochem. Cytol.*, 2(suppl.):79–83, 1956; C. Golgi, Sur la structure des cellules nerveuses, *Arch. ital. biol.*, 30:60, 1898.

Gonad

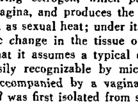
One of the paired primary sex glands—the testes in the male and the ovaries in the female. The hormones secreted by these glands, testosterone by the testes and estradiol and progesterone by the ovaries, are responsible for the development of differential secondary sex characteristics at puberty, for example, distribution of hair, skeletal differences, voice, and mammary development, as well as the more obvious influence on the primary sex organs The growth, development, and functions of female and male gonads are under the control of gonadotropic hormones produced by the anterior lobe of the pituitary. See HORMONE, ADENOHYPOPHYSEAL.

In man and most other mammals, the testes of the adult are lodged in an integumentary pouch, the scrotum, which regulates the internal temperature of the glands. The two essential components of the testis are the seminiferous tubules, which produce spermatozoa, the male sperm cells whose function is to fertilize the egg, and the interstitial cells of Leydig, which secrete the hormone testosterone. Testosterone was isolated in pure crystalline form in 1935 by E. Lacquer and his coworkers and, in the same year, proof of the structure of testosterone was accomplished by partial synthesis by A. Butenandt, L. Ruzicka, and their coworkers as Δ^4-adrosten-17β-ol-3-one:

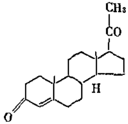


In addition to stimulating the growth and maintaining the functions of the secondary sex organs (prostate, seminal vesicles, penis, and accessory organs), testosterone acts as a potent anabolic agent. It increases the muscle mass and exercises a marked stimulus on bone growth until the epiphyses or cartilage-bone joints are closed. See ANDROGEN.

The ovaries in humans are flattened glands lying on the sides of the pelvic cavity. The cortex of the gland in sexually mature individuals contains follicles and corpora lutea, in various stages of development. The follicle is brought to maturity under the influence of one of the pituitary gonadotropic hormones, the follicle-stimulating hormone (FSH) (see PITUITARY GLAND). The mature follicle manufactures the ovarian hormone estradiol-17β, a naturally occurring estrogen, which passes to the uterus and vagina, and produces the condition of estrus known as sexual heat; under its influence a characteristic change in the tissue of the vagina occurs, so that it assumes a typical cornified appearance easily recognizable by microscopic examination, accompanied by a vaginal discharge. Estradiol-17β was first isolated from sows' ovaries by E. A. Doisy and his coworkers in 1936 and was partially synthesized in 1940 by H. H. Inhoffen and coworkers:



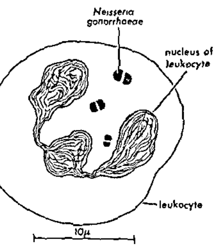
In the process known as ovulation, the mature Graafian follicles fill with fluid and move toward the surface of the ovary, where they rupture periodically, releasing the ova or female germ cells. The cells of the ruptured follicle, under the influence of another pituitary hormone, interstitial-cell-stimulating hormone or ICSH, are converted to a new structure, the corpus luteum or "yellow body." It is the mature corpus luteum that produces the hormone progesterone, which has the function of conditioning the lining of the uterus for the reception and development of the fertilized egg. The production of progesterone varies during the normal ovulatory cycle, and its output is greatly increased during pregnancy. Progesterone was first isolated from corpora lutea of pregnant sows in 1934 by A. Butenandt, K. H. Slotta, A. Wettstein, and their coworkers. In the same year, its structure was first suggested by Slotta and proved by Butenandt and his associates:



See ESTROGEN; OVARY; TESTIS. [C.H.L.]

Gonorrhea

A prevalent, contagious venereal disease of man caused by a bacterium, *Neisseria gonorrhea*, the gonococcus. This microorganism on autolysis liberates an endotoxin. The disease is contracted principally by sexual intercourse with an infected person. Nonvenereal forms of the disease are ophthalmia neonatorum, a conjunctival infection of the newborn contracted by an infant during passage through the lower birth canal of an infected



Urethral smear from gonorrhea (Gram's stain). (From W Burrows, et al., Textbook of Microbiology, 16th ed., Saunders, 1954)

mother, and gonococcal vulvovaginitis in young girls.

The incubation period of gonorrhea varies from 2 to 10 days. Early

ing sensation and frequent desire to urinate. The untreated disease may subside spontaneously and after a period of quiescence again manifest itself as an inflammation of the uterine tubes (salpingitis), an infection of the blood stream (septicemia), an inflammation of the endocardium of the heart (endocarditis), or arthritis. Inflammation of the prostate (prostatitis) and cervix (cervicitis) are frequent complications of the disease. Women are often asymptomatic carriers of the infection.

The disease is widely disseminated and the estimate of 10% morbidity in this country is considered conservative. About 1,000,000 cases of gonorrhea occur in the United States annually. The causative organism, *N. gonorrhea*, is a gram-negative, nonsporogenous, nonmotile sphere 0.6–1.0 microns in diameter. In stained preparations from pus the gonococcus is observed in the cytoplasm of polymorphonuclear leucocytes in pairs with the adjacent sides of the organisms slightly flattened like coffee beans. *N. gonorrhea* is a facultative aerobe and can best be isolated and incubated at 35–36°C on specially enriched media in an atmosphere with 10% carbon dioxide. On solid media the colonies are small, transparent, and convex. After 2 to 4 days they develop a lobate margin and are greyish-white with the appearance of pearly opalescence when viewed with transmitted light. *See* NEISSERIACEAE.

Gonococcal infection confers little or no immunity, with reinfection often occurring. Diagnosis is dependent upon identification of the gonococcus in the purulent discharges from the infected tissues and its isolation in cultures. Serologic tests are not dependable. The administration of vaccines and immune serum has little therapeutic value. The sulfonamides are often ineffective because a high percentage (80%) of the strains isolated from acute cases in recent years are sulfonamide resistant. Penicillin therapy is most effective but penicillin resistance among strains of the gonococcus has been noted with increasing frequency. Drug sensitivity among patients has increased also. Streptomycin, chloromycetin, and aureomycin may be used as alternative therapeutic agents. *See* ANTIBIOTIC; ENDOCARDITIS; GRAM'S STAIN; SEPTICEMIA; SULFA DRUGS. [C.M.C.]

Bibliography: American Public Health Association, *Diagnostic Procedures and Reagents*, 3d ed., 1950.

Goose

Any member of the subfamily Anserinae, family Anatidae. The geese are distinguished from the other waterfowl of this family primarily by their larger size and heavier bodies. They are noisy in flight and many of the species fly in V-shaped formations. Except for the brants, the geese are somewhat more terrestrial than ducks, feeding mostly on land in meadows and stubble fields. In the water they feed by tipping. Geese eat only plant food. The sexual dimorphism so characteristic of ducks is lacking in geese. In the latter the sexes look alike throughout the year. The males take part in the care of the eggs and young. There are about 25 species of geese in the world; 8 of them occur in the United States.

The Canada goose, *Branta canadensis*, is the best known of the North American geese. It has long been a harbinger of spring, whether seen in its V-shaped flights by day or noted in the night by its rhythmic honk-honk call. This goose is a favorite of the hunter. Several distinct races are recognized, including Hutchin's goose, Richardson's goose, and the lesser Canada goose. The Canada goose breeds all across northern North America and winters from Illinois southward. Great concentrations of Canada geese follow the Mississippi flyway in migration.

Two other members of the genus *Branta* in North America are called brants. They may be distinguished by the black on their breast which reaches the water line, with no white showing; white shows above the water line on the Canada goose. The brants are somewhat smaller than most Canada geese. They are seacoast geese and rarely are seen inland. The American brant, *Branta bernicla*, is common in winter along the Atlantic Coast of the United States from New Jersey to North Carolina; it breeds in the circumpolar Arctic. The black brant, *B. nigricans*, is a sea goose of the Arctic and North Pacific, nesting in the North American and Siberian Arctic and wintering in Japan, Asia, and western North America. It is rarely seen on the Atlantic Coast.

Least migratory of the geese is the emperor goose, *Philacte canagica*. This small, blue-gray goose nests in Siberia and Alaska and only rarely wanders in winter as far south as northern Cali-

fornia. The white-fronted goose, *Anser albifrons*, is a Holarctic species, breeding in the Arctic and subArctic and wintering in Texas and Mexico. It is seldom seen east of the Mississippi River.

The blue goose, *Chen caerulescens*, and the snow goose, *C. hyperborea*, are often found in the same flock, and apparently interbreed rather freely. Some believe them to be races of the same species, and there is evidence that the blue form is increasing at the expense of the white bird. Except for stragglers, all the blue geese appear to winter along the Louisiana and Texas coast. *Chen rossi*, Ross's goose, is a western bird, breeding in the Arctic and wintering in interior California. It is quite similar to the snow goose. [J.D.B.]

For a discussion of domesticated barnyard geese see POULTRY PRODUCTION.

Gooseberry

A small fruit represented by about six species of the genus *Ribes* of the plant order Rosales. The gooseberry is a thorny, spreading bush which grows to a height of about 3 ft and produces red,

Gooseberry branch bearing leaves and fruits. (USDA)

yellow, or green berries. The most desirable hardier types in the United States are of American parentage, or are hybrids between American and European species. Commercial culture is limited to a few states, notably Oregon, Michigan, and Washington, but gooseberries are common in home gardens throughout most of the United States except the far South and Southwest. In 1954 the value of the gooseberry crop in the United States was $176,116.

The fruit is very acid and only a few European varieties, when fully ripe, are suitable for eating fresh. The fruit may be canned or frozen for use in pies or as preserves.

For diseases of gooseberry see CURRANT; see also FRUIT GROWING (SMALL). [J.H.CE.]

Gopher

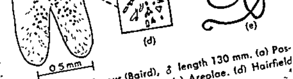
A name applied in different localities to various animals that dig. The name, however, is properly limited to the North American pocket gophers, a group of fossorial rodents comprising the family Geomyidae, common in the western and west central states and also found on the Gulf Coastal Plain.

The pocket gopher, *Geomys bursarius*; length to 1 ft. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Pocket gophers reveal their presence by large mounds of dirt which they pile at frequent intervals along their burrows. They eat plant roots and underground stems. They are capable of doing considerable damage to grasslands and alfalfa fields, but their burrows are of some value in aerating the soil and in admitting water into the ground. See RODENTIA. [J.D.B.]

Gordioidea

An order of worms belonging to the class Nematomorpha. The gordioids include all the fresh-water species of nematomorph worms, and can be distinguished by the one ventral epidermal chord, the body cavity which is filled with tissue, and the paired gonads with lateral diverticula in which the eggs ripen. The gordioids are divisible into two families, the Gordiidae that have a smooth cuticle,



Gordionus violaceus (Baird), ♂ length 130 mm. (a) Posterior end. (b) Anterior end. (c) Areolae. (d) Hairfield enlarged. (e) General appearance.

and the Chordodidae that have a rough cuticle containing conspicuous thickenings called areoles. The gordiids comprise one genus, *Gordius*. The chordodids have several genera which include *Chordodes*, *Paragordius*, *Parachordodes* and *Gordionus*. The arrangement and shape of the areoles form the most important taxonomic characteristics. In the parasitic phase each species of gordioids has a special host. This is either an aquatic or a terrestrial insect. Among the terrestrial insects, grasshoppers and carabid beetles are preferred, but it is not known how the infection takes place. Somehow the infection causes the host to seek moist or aquatic habitats where the adult gordioid creeps out to propagate. See NEMATOMORPHA. [B.J.MU.]

Gorgonacea

An order of the subclass Alcyonaria. The Gorgonacea are the horny corals which often form fanlike or featherlike colonies with branches spread radially or oppositely in one plane. They attach to objects by somewhat enlarged bases or tufts of stolons. They are more widely distributed than the

Fig. 1. (a) Skeleton of *Corallium konojoi*. (b) *Melitodes* sp. (preserved specimen). (c) Colony of *Anthoplexaura dimorpha* (preserved specimen).

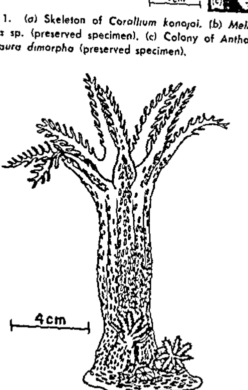


Fig. 2. Primary polyp of *Anthoplexaura dimorpha* (after K. Kinoshita).

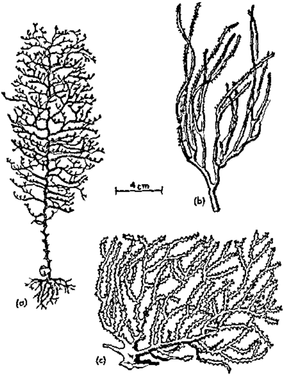


Fig. 3. (a) *Chrysogorgia flexilis* var. *africana* Kükenthal (after W. Kükenthal). (b) *Lophogorgia crista* (after W. Kükenthal). (c) *Paramuricea Kükenthali* Broch (after H. Broch).

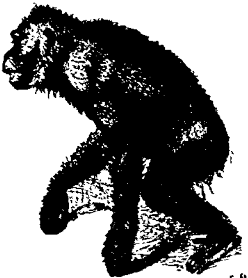
Alcyonacea and extend from the littoral zone to some great depth.

The gorgonians are characterized by the presence of an axial rod composed of horny matter or gorgonin. The order is divided into two suborders, the Scleraxonia and Holaxonia, according to the structure of the rod which has an outer cortex and an inner core, or medulla. In the Scleraxonia, which includes such genera as *Paragorgia*, *Corallium* (Fig. 1*a*,*b*), and *Melitodes*, the axial skeleton contains calcareous spicules. In the Holaxonia, which are the typical gorgonians, a spiculeless skeleton is formed by the gorgonin and a small amount of calcareous matter. The latter group includes many genera: *Gorgonia* or the common sea fan, *Paramuricea*, *Lophogorgia* (Fig. 3*b*,*c*), *Muricella*, *Anthoplexaura* (Figs. 1c, 2), *Chrysogorgia* (Fig. 3*a*), and *Keratoisis*. The last has a peculiar segmented axis with alternating calcareous and horny portions.

Most gorgonians are monomorphic. The autozooid resembles that of the Alcyonacea in many respects. Corallidae and *Paragorgia* are dimorphic; the siphonozooid has no tentacle. See ALCYONARIA. [K.A.]

Gorilla

Either of two species of African apes of the genus *Gorilla*, family Pongidae. These are the largest of the primates; a large male may stand 6 ft tall and weigh as much as 600 lb; females are much smaller



The gorilla, Gorilla savagei; height, male 5 ft 7½ in (From P. Martin Duncan, ed., Cassell's Natural History, Cassell & Company, Ltd., London)

They are covered with long, black hair and are tailless. Gorillas live in family troops of one male accompanied by several females and their young. They are powerful and dangerous animals, but are peaceful unless molested.

Young gorillas are more apt to walk upright than the older animals which usually walk on all four legs. Gorillas are wandering animals; the young ones often build a platform of sticks in a tree on which to rest at night and move on to a new locality the following day. Old males sleep on the ground. Gorillas are vegetarians. See PRIMATES. [J.D.B.]

Gout

An acute or chronic disease of protein metabolism marked by recurrent arthritis. In this disease there is a derangement of purine metabolism so that increased amounts of uric acid are present. The increase is thought to be caused by increased production or decreased excretion of uric acid from the kidney, or by both. See PROTEIN METABOLISM.

Uric acid is a normal breakdown product of purine metabolism and is derived from either ingested foods or body tissues.

Gout occurs most frequently in middle-aged males and is passed as a familial or hereditary trait which for some unknown reason does not appear as often in females.

The uric acid becomes deposited as urates in soft tissues, especially around the joints, in the cartilages of the ear, along the shafts of long bones, in the kidney, and occasionally, on the heart valve. Such deposits, when superficial, can be seen grossly as reddish, inflamed masses called tophi. The inner portion consists of chalky-white urate deposits and a surrounding inflammatory reaction.

Apparently the mere presence of a tophus does not produce the arthritis and other pain. When the disease is in remission, the tophi are still present. Recently drugs have appeared which greatly aid in uric acid excretion. The clinical course is marked by long intervals of almost complete remission, but the disease will flare up from time to time as the result of stress, infection, or unknown causes which may include temporary endocrine imbalances associated with certain adrenal cortex hormones. [E.G.ST.]

Governor

A device used to control the speed of a prime mover. A governor protects the prime mover from overspeed and keeps the prime mover speed at or near the desired revolutions per minute. When a prime mover drives an alternator supplying electrical power at a given frequency, such as 60 cycles per second, a governor must be used to hold the prime mover at a speed that will yield this frequency. An unloaded diesel engine will fly to pieces unless it is under governor control.

Speed control. A governor regulates the speed of a prime mover by properly varying the flow of energy to or from it. In the case of gas and steam turbines and internal combustion engines, the fuel furnishes the energy to the prime mover. For such applications, the governor usually controls the speed of the unit by regulating the rate at which fuel, and hence energy, is furnished to the prime mover. The governor controls the fuel flow so that the speed of the prime mover remains constant regardless of load and other disturbances, or changes in accordance with such operating conditions as changes in speed setting.

For a diesel engine the governor is connected to the rack which controls the amount of fuel injected. A governor on a gas or gasoline engine is attached to the engine throttle. A steam turbine governor strokes the steam valve or valves which regulate the steam flow to the turbine.

The output mechanism of a gas turbine governor is connected to the fuel valve with the stroke normally limited in each direction by the allowable combustion chamber temperature and other factors. If the fuel rate is too low, the fire will go out. Compressor stall must be avoided. To give rapid control, combustion chamber temperature is often computed in the governor by measuring other variables and applying the laws of thermodynamics.

A hydraulic turbine governor regulates the flow of water to the turbine by varying the openings of gates or other components.

An aircraft propeller governor varies the pitch of the propeller to keep constant the speed of the engine attached to the propeller. This type of governor varies the load on the engine and thus controls the speed by regulating the energy flow from the prime mover.

Ballhead governor. The speed of a prime mover is usually measured by a ballhead that contains flyweights driven at a speed proportional to the speed of the prime mover. The force from the

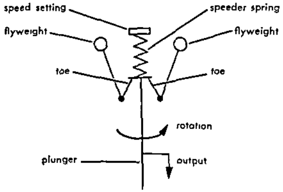


Fig 1. Ballhead of governor.

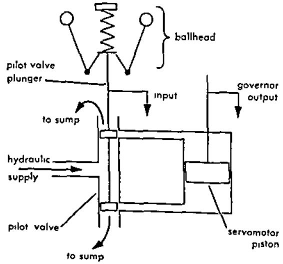


Fig. 2. Isochronous governor.

weights is balanced, at least in part, by the force of compression of a speeder spring (Fig. 1). The upper end of this spring is positioned according to the speed setting of the governor.

The ballhead toes press against one end of a plunger. In the simplest version of a governor, the plunger position is a function of the engine speed as a result of the balance between the centrifugal and spring forces. The plunger is directly connected to the throttle or other energy controlling means of the prime mover. Because the governor output power is drawn directly from the speed measuring means, the output power and the precision of such a governor are severely limited.

Governor response. In automatic control theory, the input to the governor is taken to be the difference between the speed setting and prime mover speed. This difference is the speed error. For the simplest governor, the position of the plunger is the output. This governor is proportional since the governor output is proportional to the governor input. In equilibrium, where the engine is running in steady state, the speed of the prime mover depends on both the load and the speed setting. With the mechanical governor and a fixed speed setting, the equilibrium speed decreases as the load increases. A governor with this property is a droop governor, or a governor operating on droop. The governor-prime mover unit is then said to be running on droop.

To increase the power output of a governor, a hydraulic amplifier is often employed. A governor that keeps the speed of a prime mover constant is said to be isochronous. In a simple isochronous governor, the ballhead senses the speed and strokes a pilot valve plunger that regulates the flow of fluid to a servomotor (Fig. 2). Normally, the fluid is oil. The servomotor is a piston in a cylinder; the piston is connected to the engine throttle or equivalent energy controlling mechanism. The position of the servomotor piston is the output of the governor. This output is made proportional to the integral of the governor input (speed error), and the governor is therefore an integral, or reset, controller. This governor is intended to bring the prime mover speed back to the speed setting after any change in load.

A hydraulic droop governor is obtained from the simple isochronous governor by the introduction of feedback from the governor output to the pilot valve (Fig. 3). This governor behaves like a simple mechanical governor except that a smaller ballhead is generally employed, greatly improving the precision of the governor by reducing hysteresis, dead band, and friction in the ballhead. The power output of the governor is much larger so that the effect of the load on governor performance incurred in moving the throttle or equivalent mechanism, is considerably diminished. A hydraulic governor may be sensitive to speed changes of as little as 1/1000 of 1%. For normal disturbances, prime mover speed error may be kept to 0.1% or better. The power output of such a governor may exceed 1 horsepower. The mechanical version may be insensitive to speed changes of 1%; the output is generally limited to 1/50 horsepower (hp). The lag in the hydraulic governor is usually below 1/10 sec, whereas a lag of ½ sec or more is common for the mechanical governor.

Use of dashpot. The performance of the simple isochronous governor is often greatly improved by

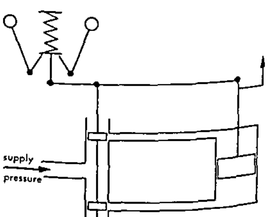


Fig. 3 Hydraulic droop governor.

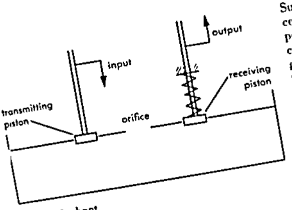


Fig. 4. Dashpot.

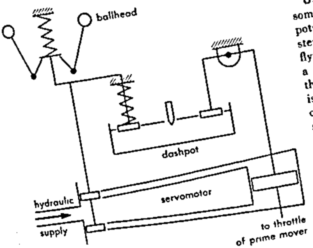


Fig. 5. Isochronous dashpot governor.

the introduction of a dashpot in the feedback path from the output to the ballhead. If there is little damping in the prime mover, instability often occurs when the simple isochronous governor is used, whereas this instability is removed when the dashpot is incorporated. A system is stable when for each disturbance that dies out, response of the system settles to an equilibrium condition. When instability occurs, the prime mover speed oscillates continuously or increases indefinitely until the unit flies to pieces.

In a dashpot (Fig. 4) deflection of a spring-loaded piston from equilibrium depends on the velocity of the other, or transmitting piston. In its simplest version this dependence is a proportionality. The spring-loaded piston is termed the receiving piston. The position of the transmitting piston is the dashpot input, whereas the position of the receiving piston is the output. Between input and output there is a time lag. The receiving piston attains its equilibrium when the pressure drop across the dashpot orifice is zero. Oil is used as the dashpot fluid.

When a dashpot is incorporated into a governor (Fig. 5), the governor output becomes a function of the speed error and the integral of this error. Such a governor is a proportional plus integral controller. The velocity of the servomotor then depends on prime mover speed and prime mover acceleration. The time lag in the dashpot makes the governor sensitive to prime mover acceleration. As this lag is increased, the response of the governor to prime mover acceleration tends to increase. This increase in lag is accomplished by moving the dashpot needle valve, which controls the orifice area, toward the closed position.

Instead of mechanical feedback from the governor, force feedback is generally preferred (Fig. 6). An isochronous dashpot governor is turned into a droop governor by adding direct mechanical feedback from the servomotor to the ballhead.

Use of flywheel. Acceleration governors are sometimes used in place of governors with dashpots. In such governors a flywheel is employed instead of a dashpot. The prime mover drives the flywheel through a spring. This combination yields a motion proportional, except for a time lag, to the acceleration of the prime mover. This motion is added mechanically or otherwise to the output of the governor ballhead, and the result is used to stroke the pilot valve plunger. The pilot valve meters the flow of fluid to the servomotor cylinder. The force from the flywheel is usually amplified hydraulically before the motion is added to the ballhead output. The outputs of acceleration governors tend to jiggle because they are more sensitive to high-frequency variations in the governor ballhead drive.

Governor applications. Governors for large hydraulic turbines require a second stage of hydraulic amplification. The governor servomotor piston is connected to the plunger of a relay valve that regulates the flow of oil to the turbine servomotors. The governor servomotor is then termed the con-

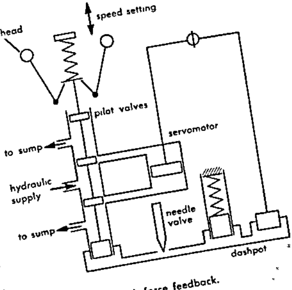


Fig. 6. Governor with force feedback.

trollet. Turbine servomotors often require 10 hp or more.

If at all times the driving torque is equal to the load torque, the prime mover will neither accelerate nor decelerate. This principle is employed in load control governors where a measurement of load is employed to rapidly position a load servomotor piston so that the driving and load torques are approximately equal, and speed errors kept small. To the motion of this piston is added the motion of the regular governor piston, which now functions as a vernier control.

The speed setting of a governor is often adjusted by an electric, hydraulic, or pneumatic motor as a function of auxiliary variables. Thus in an electrical power system the deviation of the frequency of the system from the desired value is used to position the governor speed setting so as to bring the frequency to the right value. Adjustment of governor speed settings is required to keep electric clocks on time. Oil and gas pipeline governors control pressure rather than speed. The pressure in the line is measured and the speed setting of the governor is adjusted accordingly. This affects the speed of the prime mover, which drives a compressor, raising or lowering the pressure as required.

Paralleled prime movers. Prime movers may be paralleled to supply power to the same load. In an electrical power system, prime movers drive alternators electrically connected to the system. At most, one of the governors of paralleled prime movers can be isochronous; the rest must be on droop. The prime movers may all be on droop. Rated load on a prime mover is 100% load; rated speed is 100% speed. If as the load is increased 100%, the prime mover speed falls u% the governor is on u% droop. The amount of droop is determined by the amount of servomotor motion fed back to the governor ballhead. In operation the percentage of droop is usually fixed, while the speed setting is adjusted to make the prime mover take its assigned load. A droop of 5% is common. The droop generally falls in the range of 1–10%.

When two or more identical prime mover-generator units are paralleled, and one is controlled by an isochronous governor while the others are on droop, electrical coupling will force all units to run at the same steady-state speed. The alternators then supply electrical power at the same frequency. The load taken on by a unit with a droop governor is now a function of the speed setting of the governor (Fig. 7). At 100% speed, the speed of the prime movers produces the required system frequency. The load taken on by a unit is determined by the intersection of its load line with the 100% speed line. As the speed setting N_s of the unit is increased, the load line moves up, retaining the same slope but increasing the load taken on by the unit. By adjusting the speed settings of the droop governors, load may be arbitrarily divided among the generating units.

Aircraft engines are synchronized and synchrophased by using one engine as the master, and adjusting the speed settings of the governors on the other engines to make them follow the master.

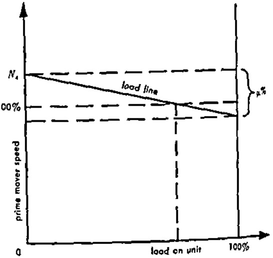


Fig. 7. Droop characteristic.

Governor design. Proper design of a governor involves making the governor characteristics fit those of the prime mover and load so as to give acceptable over-all performance. Desirable and undesirable nonlinearities occur in governors to complicate design calculations. Thus centrifugal force varies as the square of speed. To compensate for this, nonlinear speeder springs are often employed. Bypass is often used in dashpots to limit the range of the dashpot output. When the receiving piston for a dashpot (Fig. 4) is sufficiently far removed from the equilibrium position, passages are uncovered, which permit the fluid to flow from one side of the piston to the other, thus limiting its travel.

The output of a generator driven by the prime mover is sometimes used to obtain a voltage proportional to prime mover speed. This voltage, properly filtered, is fed to an electronic, magnetic, or other circuit, and eventually to a transducer to move the pilot valve. The output pressure of a hydraulic pump is sometimes employed as a measure of prime mover speed. Other speed-measuring means are utilized.

Limit or topping governors are often employed to meet fail-safe specifications. The limit governor is usually a mechanical governor that takes over control from the main governor to shut the prime mover down if its speed reaches a fixed overspeed above its rated speed. The flyweight force increases with an increase in the distance of the flyweights from the axis of rotation of the ballhead; this flyweight force opposes the speeder spring force. The flyweight force increases as the square of prime mover speed. When it is sufficiently large, the governor gain becomes infinite, so that a small input to the governor causes an unlimited output. This causes a snap action of the governor at the overspeed for which the governor is set.

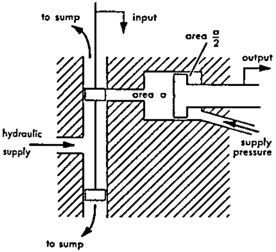


Fig. 8. Two-way valve—differential servomotor.

The natural frequency of a ballhead is generally so high that the ballhead lag can be neglected in first approximation studies of governor-prime mover system performance. When the ballhead drive involves dominant frequencies near the ballhead resonant frequency, damping is sometimes provided in the ballhead to prevent excessive governor output jiggle.

In the case of a hydraulic governor, the speeder spring force is aided by the hydraulic reaction force tending to close the pilot valve. This reaction force occurs because of a change in momentum of the fluid passing through the valve and is proportional to the opening of the valve for small valve openings. This force behaves like the force of a linear spring. The effect of hydraulic reaction is to diminish the gain of the ballhead, so that a given input causes a smaller output.

In the interests of economy the four-way valve servomotor (Fig. 2) is often replaced by a two-way valve-differential servomotor (Fig. 8). The area on one side of the servomotor is half the area on the other side. The supply pressure on the small area is sometimes replaced by a spring.

Governors are normally designed so that the response to a sudden disturbance dies out in 1–20 sec.

A load is termed isolated if the power supplying this load comes from one prime mover only. In the design of a governor for a prime mover supplying power to an isolated load, a prime mover differential equation is obtained relating the input to the prime mover, such as throttle position, to the speed of the prime mover, which is taken to be its output. This equation is obtained by mathematically balancing the torques on the prime mover shaft. The equation also involves a delay between a change in the input to the prime mover and the resulting change in the driving output torque. Part of the delay is a dead time. This dead time is normally between 1/100 and 1/2 sec. The rest of the delay is usually in the order of 1/10 sec.

The nature of a prime mover may introduce other terms in the equation. For example, with hydraulic turbines the equation is complicated because of water hammer; a sudden closing of the gates to decrease the speed of the turbine causes the turbine to speed up initially, resulting in a correction opposite to that desired. As the gates are moved, pressure waves travel up and down the hydraulic conduit from the source of the water to the discharge from the turbine. These waves tend to unstabilize the unit. *See* WATER HAMMER.

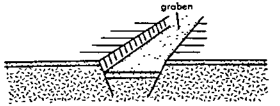
The equation of the prime mover is combined with the equation of the governor to yield an equation for over-all system response from which the governor response is determined. When full load on a diesel engine is rejected, an overspeed of 3% is often considered acceptable with the response dying out in 1 sec or less. For a hydraulic turbine this overspeed may be 30–40% and the response may endure as much as 10–20 sec. Performance for other prime movers under governor control falls between these entremes. [R.O.]

Grab bucket

A digging and materials-handling tool with hinged sides or leaves. It is particularly suited for digging deep holes or for raising material to high places. A grab bucket may have 2, 3, or 4 movable sides. A 2-sided bucket is somewhat rectangular in shape and is called a clamshell. A 3- or 4-leafed bucket is ball-shaped and is called an orange peel. For hard digging, clamshells are equipped with hardened prongs or teeth; for handling loose material such as sand, cinders, or fine coal, no teeth are used. Standard clamshell capacities range from 3/8 yd^3 to 6 yd^3 but larger sizes have been made. Orange peel buckets are manufactured with capacities up to 3 yd^3, but this size is unusual. Most orange peels are small; some are only slightly bigger than a large grapefruit. They are used mainly for cleaning sewer catch basins, vertical pipes, caissons, and other small vertical openings. [E.M.Y.]

Graben

A block of the earth's crust, generally with a length much greater than its width, that has been dropped relative to the blocks on either side (see illustration). The size of a graben may vary; it may be only a few inches long or it may be hundreds of miles in length. The faults that separate a graben from the adjacent rocks are inclined from 50–70° toward the downthrown block and have displacements

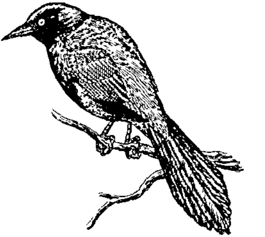


Simple graben. (*From A. K. Lobeck, Geomorphology, McGraw-Hill, 1939*)

ranging from inches to thousands of feet. The direction of slip on these indicates that they are gravity faults. Graben are found in regions where the crust has undergone extension. They may form in the crests of anticlines or domes, or may be related to broad regional warpings. *See* FAULT AND FAULT STRUCTURES; HORST; RIFT VALLEY; WARPING, EARTH CRUST. [P.H.O.]

Grackle

Either of two large United States blackbirds with long tails. *Quiscalus quiscula* is the more widely distributed. The Atlantic Coast race of this bird is known as the purple grackle; the subspecies of the interior states is called the bronzed grackle. This is a thriving species, very common in grain belt towns where it roosts in town and flies out to the grain fields to feed. It is a pest in many agricultural



The bronzed grackle, *Quiscalus quiscula*; length to 14 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

areas, especially in late summer when it congregates in large flocks. The larger boat-tailed grackle, *Cassidix mexicanus*, which may grow to 17 in. in length, is characterized by a huge, keel-shaped tail. It is found from Virginia southward into Central America. There are two related tropical species. All of them may be called crow blackbirds. *See* CROW; PASSERIFORMES. [J.D.B.]

Gradient of a scalar

The result of the application of the distributive vector differential operator ∇, $\nabla = \mathbf{i}\,\partial/\partial x + \mathbf{j}\,\partial/\partial y + \mathbf{k}\,\partial/\partial z$, to a differentiable scalar function $S(x,y,z)$; thus $\nabla S = \mathbf{i}\,\partial S/\partial x + \mathbf{j}\,\partial S/\partial y + \mathbf{k}\,\partial S/\partial z$. The letters $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, are symbols for the base vectors associated with x, y, z. The gradient of S is also denoted by grad S, while the symbol ∇ is usually called del and less frequently nabla.

The key properties of ∇S are [illegible]

[illegible] temperature at each point of a *room*, then from a given point of observation S may increase in certain directions (say toward a heat source) and decrease in others. Specifically, let P be a given point, C a curve given in terms of its arc length s by $\mathbf{r} = x(s)\mathbf{i} + y(s)\mathbf{j} + z(s)\mathbf{k}$ and further, let C pass through the point P and satisfy the requirement $d\mathbf{r}/ds|$ at $P = \boldsymbol{\tau}$, for $\boldsymbol{\tau}$, a given prescribed unit vector. Along C, $x = x(s)$, $y = y(s)$, $z = z(s)$ and $S = S[x(s), y(s), z(s)]$ and if the functions are of class C'; then

$$\left.\frac{dS}{ds}\right|_{P,C} = \left.\frac{\partial S}{\partial x}\frac{dx}{ds} + \frac{\partial S}{\partial y}\frac{dy}{ds} + \frac{\partial S}{\partial z}\frac{dz}{ds}\right|_{P}$$
$$= \nabla S|_P \cdot \boldsymbol{\tau} = |\nabla S|_P| \cos(\nabla S|_P, \boldsymbol{\tau})$$

Here P, C indicates that the derivative is taken at the point P and for the curve C. This equation shows that $dS/ds|_{P,C}$ depends on S, P, and $\boldsymbol{\tau}$ only; hence, the alternative notation $dS/ds|_{P,\tau}$ is more appropriate. The quantity $dS/ds|_{P,\tau}$ is called the directional derivative.

Evidently, if (1) S and P are fixed, (2) $\boldsymbol{\tau}$ is variable in direction, and (3) $\nabla S|_P \neq 0$ then $dS/ds|_{P,\tau}$ can vary with the direction of $\boldsymbol{\tau}$ and will take on its maximum value when $\cos(\nabla S|_{P,\tau}) = 1$. Thus $\nabla S|_P$ points in the direction of the greatest rate of increase of S at P and $|\nabla S|_P|$ is this greatest rate of increase. Also, if $S(x,y,z) = S(a,b,c)$, $(a,b,c$ fixed) defines a surface then by confining C to the surface it may be shown that if $\nabla S|_P \neq 0$, then $\nabla S|_P$ is perpendicular to the surface. *See* CALCULUS OF VECTORS. [H.V.C.]

Gradient wind

A hypothetical wind based upon the assumption that the sum of the horizontal components of the Coriolis force and the atmospheric pressure gradient force per unit mass is equivalent to a wind acceleration which is normal to the direction of the wind itself (centripetal acceleration), with the implication that there are no viscous forces acting. The direction of the gradient wind is the same as that of the geostrophic wind. Its speed is given by

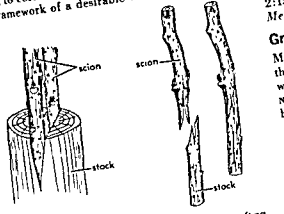
$$V_{\text{grad}} = \frac{V_{\text{geo}}}{\tfrac{1}{2} + \sqrt{\tfrac{1}{4} + \dfrac{V_{\text{geo}}}{2R\Omega \sin\phi}}}$$

where V_{geo} is the speed of the geostrophic wind, Ω is the angular speed of rotation of the earth, ϕ is the latitude, and R is the radius of curvature of the air trajectory, considered positive when the trajectory is curved in the cyclonic sense (center of curvature on low-pressure side) and negative when the curvature is anticyclonic. The gradient wind speed is less than the geostrophic speed when the air moves in a cyclonically curved path and greater when the air moves in an anticyclonically curved path. The gradient wind is a good approximation of the actual wind and is often superior to the geostrophic wind, particularly when the flow is strongly curved in the cyclonic sense. *See* CORIOLIS ACCELERATION AND FORCE; GEOSTROPHIC WIND; STORM; WIND. [F.S.]

Grafting of plants

Vegetative propagation by joining a scion (a short section of the stem) to an understock in such manner that the two grow together and continue development as a single plant. Success depends upon the scion being nearly dormant, as in spring or early summer, plus genetic compatibility of scion and stock, firm union of the cambia of these two (*see* MERISTEM, LATERAL), and protection from desiccation. The various kinds of grafts may be included under either topworking or repair grafting.

Topworking is employed to change the variety of fruits, to correct pollination problems, or to change the framework of a desirable variety which is subject to trunk or crotch injury. Topworking includes three methods: root grafting, in which a scion 3–6 in. long is grafted on a root; crown grafting, a similar scion graft made at the root crown just below ground level; and top grafting, including top budding, cleft grafting, and frameworking.



Two common types of grafting. (From H. J. Fuller and O. Tippo, *College Botany*, rev. ed., Holt, 1954)

Repair grafting is of two kinds: bridge grafting, in which each of several scions is grafted in two positions on the stock, one above and the other below an injury; and inarching, in which two plants growing on their own roots are grafted together and one plant severed from its roots after the graft union is established. Other methods of grafting are whip or tongue, cleft, bark, saw-kerf, stub or notch, side, and veneer. *See* BUDDING; PLANT ANATOMY; REPRODUCTION, PLANT. [J.E.GU.]

Grain boundaries (metallurgy)

The surfaces of contact between two perfect pieces of crystal. When a crystal is grown from a melt without any particular care taken in the process, the result is usually a crystal made up of various sized crystallites called a polycrystal. As a result of accidents of growth, a nucleus of crystal will grow until it meets another growing crystallite, with a grain boundary forming between the two randomly oriented crystals. The crystallites are also called grains; hence the name, grain boundary. The grain boundary itself is a highly disordered region if the misorientation of the two crystallites is large, and it has some of the characteristics of a liquid. For instance, diffusion takes place very rapidly along it. On the other hand, a polycrystal is much stronger than a single crystal because the grain boundaries in general impede the passage of dislocations. Grain boundaries, as well as dislocations, are considered to be types of crystal defects. For a discussion of dislocations, *see* CRYSTAL DEFECTS; *see also* CRYSTAL GROWTH; TWINNING (CRYSTALLOGRAPHY). [R.M.T.]

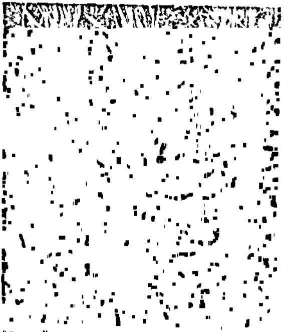
Bibliography: R. W. Cahn et al., Symposium on polygonization in crystals, *Prog. in Metal Phys.*, 2:151–202, 1950; D. McLean, *Grain Boundaries in Metals*, 1957.

Grain crops

Members of the grass family, grown primarily for their edible seeds, including the small grains wheat, rye, oats, barley, and rice (*see* GRAMINALES). Samples of grain crops not considered to be small grains are corn and grain sorghum.

The acreage planted to grain crops in different areas of the United States has varied greatly in the agricultural history of this country. Even in recent times acreage adjustments have occurred because of changes in market demands, competition from other crops, changes in incidence of destructive diseases, and development of new varieties. For example, introduction of winter-hardy wheat varieties from Russia made it possible to expand winter wheat production greatly in the Great Plains states and subsequently, by further breeding, made possible the successful culture of winter wheat as far north as central Minnesota. *See* BREEDING (PLANT). Development of higher-yielding varieties and improved cultural practices, coupled with less per capita use, have resulted in a gradual but consistent reduction of total wheat acreage in the United States. Barley, once an important crop in Michigan, Wisconsin, Minnesota, and Iowa, has now moved from these states westward to a small area in the Red River Valley in North Dakota and Minnesota and to the dryland farming areas of California, Montana, and Colorado. This change occurred largely because destructive diseases made barley yield and quality less certain than those of other small grains.

Often such grains as oats are grown primarily as a companion crop for the establishment of forage-crop seedings (see illustration). In these cases the net cash returns, in comparison with those from other competing crops, may be of secondary importance. In multiple-crop areas, a more even distribution of labor through the crop-growing season may favor including small grains in the crop rotation. In other areas, such as the Red River Valley, western Kansas, Nebraska, and Oklahoma, small



An excellent stand of alfalfa established with oats as a companion crop. (*Photograph from J. D. Allen and Son*)

grains generally produce greater net income than alternative crops because of lower production costs.

The harvested acreage, yield, and average farm price per bushel of five small grains for the period 1952–1954 are shown in the table.

Crop	Acreage	Yield per acre, bushels	Price per bushel, dollars
Wheat	61,099,700	18 2	2 07
Oats	39,930,300	33 1	71
Barley	9,941,300	28 0	1 19
Rye	1,495,000	12 8	1 42
Rice	2,166,300	2455 cwt	5 21 per cwt

See BARLEY; CORN; OATS; RICE; RYE; SORGHUM; WHEAT.

[I.J.J.]

Gram

The unit of mass in the centimeter-gram-second system of units. It is defined as being equal to 0.001 kilogram. The gram is abbreviated g or gm (g in this encyclopedia). A gram weighs 15.43236 grains or 0.03215 *ounces troy*. There are 1000 milligrams in 1 gram. *See* KILOGRAM; MASS; METRIC SYSTEM; UNITS, SYSTEMS OF.

Gram's stain

A differential stain used widely in bacteriology as one of the criteria in the identification and classification of bacteria. The Gram stain reaction is based on differences among stained bacteria in their resistance to decolorization by neutral solvents such as alcohol or acetone. The fixed smear is stained with a slightly alkaline solution of a suitable basic dye, treated with a mordant and then with a neutral decolorizing agent, and usually counterstained with a dye solution of a contrasting color. Cells or structures that hold the color of the primary stain are called gram-positive and those that take the color of the counterstain, gram negative. *See* BACTERIA, TAXONOMY OF.

A marked correlation exists between the gram reaction and the susceptibility of bacteria to the bacteriostatic action of dyes and antibiotics, and to the action of physical agents. For example, penicillin is active against gram-positive bacteria, whereas streptomycin is active for the most part against gram-negative bacteria.

Smears may be fixed by heat or chemically. The most suitable dyes are the methyl violets, and the most satisfactory mordants, iodine and picric acid. Depending on the method, the differentiating agent may be 95% ethanol, acetone, or chloroform. Mixtures of alcohol and acetone are preferred by some workers. The counterstain should produce maximum contrast with the primary stain and should not be sufficiently strong or concentrated to replace the primary stain.

There are many versions of the Gram stain. Some of these versions were developed for specific purposes. A recommended procedure is to (1) stain the heat-fixed smear for 1 min with Hucker's crystal violet solution composed of 2 g crystal violet, some 95% ethyl alcohol, 0.8 g ammonium oxalate, and 80 ml of distilled water, (2) rinse with tap water not more than 2 sec, (3) treat for 1 min with an iodine solution composed of 1 g of iodine and [illegible]

smear dry again, (8) counterstain with a safranine solution composed of 10 ml of a 2.5% solution in 95% ethanol + 100 ml of distilled water, (9) rinse, (10) air-dry, and examine. Gram-positive substrates appear dark violet, and gram-negative ones, red.

The Gram reaction may be affected by variation in technique; therefore, it is necessary to name or describe the procedure used and to adhere rigidly to its recommended details. The Gram reaction may also be affected by age of the culture, composition of the medium, autolysis, exposure to ultraviolet light and animal sera, and so forth.

Many theories have been proposed to explain the mechanism of the Gram stain. For a while it was believed that gram-positive bacteria owed this property to a special magnesium-ribonucleoprotein-carbohydrate complex that forms an alcohol-insoluble compound with the dye and iodine. This was supported by the fact that gram-positive bacteria may be rendered gram-negative by treatment with ribonuclease, and that the cells so obtained could again be transformed to gram-positive ones by covering the bacteria with extracted magnesium-ribonucleoprotein. Recent studies have, however, focused attention on a particular lipopolyglycerophosphate-protein complex, found in the cell envelopes of all gram-positive but not of gram-

negative bacteria, as the material that is directly responsible for the outcome of the staining procedure. [G.K.N.]

Gram-equivalent weight

The equivalent weight of an element or compound expressed in grams (g), that is, the equivalent weight on a scale on which the equivalent weight of oxygen is 8.000 g. The gram-equivalent weight of an electrolyte is usually the weight in grams which is associated with 1 faraday of electricity. For example, the gram-equivalent weight of sodium chloride, $Na^{+}Cl^{-}$, is equal to the gram-molecular weight; that of calcium chloride, $Ca^{++}Cl_2^{-}$, is one-half the gram-molecular weight; that of aluminum sulfate, $Al_2^{+++}(SO_4^{--})_3$, is one sixth of the gram-molecular weight. See ELECTROCHEMICAL EQUIVALENT; ELECTROLYSIS; ELECTROLYTE; EQUIVALENT WEIGHT; GRAM-MOLECULAR WEIGHT. [T.C.W.]

Graminales

A large and important order of the plant subclass Monocotyledoneae, including two families: the grasses (*Gramineae*) with 500 genera and about 4500 species, and the sedges (*Cyperaceae*) with 85 genera and over 3000 species. The grasses usually have hollow stems and linear, two-ranked leaves consisting in most cases of an open sheath enveloping the stem, a blade, and an intervening appendage (ligule). Flowers are in spikelets and the fruit usually is a caryopsis (grain). This is the most widely distributed and economically the most important family of vascular plants. Grasses supply food for man and other animals, they are used medicinally, and they furnish materials for matting, upholstery, bedding, and the making of hats and paper. See separate articles describing wheat, rice, corn, and other important grasses.

The sedges have a grasslike habit, a solid, usually triangular stem and three-ranked leaves with a closed sheath, narrow blade, and no ligule. Flowers are borne in spikelets and the fruit is an achene or nutlet. They have a world-wide distribution but little economic importance. See BAMBOO; CITRONELLA; see also EMBRYOPHYTA; GRASS CROPS; MONOCOTYLEDONEAE; PLANT KINGDOM. [P.D.S.]

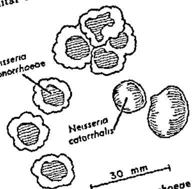
Gram-molecular weight

The molecular weight of an element or compound expressed in grams (g), that is, the molecular weight on the scale on which the atomic weight of oxygen is 16.000 g. The ratio of the gram-molecular weights of any two elements or compounds must be identical with the ratio of the absolute weights of their individual molecules. Therefore the gram-molecular weights of all elements or compounds contain the same number of molecules. This number, called the Avogadro number N, is 6.023×10^{23}.

Since they contain the same number of molecules, the gram-molecular weights of all gases occupy the same volume at the same temperature and pressure. At 0°C and 1 atmosphere pressure this volume, called the gram-molecular volume, is 22.4 liters. See AVOGADRO NUMBER; GAS; MOLAR VOLUME; MOLECULAR WEIGHT. [T.C.W.]

Gram-negative diplococci

Bacteria of the genus *Neisseria*. The family Neisseriaceae includes two important pathogens, *N. gonorrhea* and *N. meningitidis*, the causative agents of gonorrhea and epidemic cerebrospinal meningitis respectively. *N. flavescens*, a chromogenic diplococcus, may also cause meningeal infection. All members of the genus are parasitic. *N. sicca* and *N. catarrhalis* are nonpathogenic inhabitants of the mouth and of the upper respiratory and genital tracts in man. Other gram-negative diplococci are species in *Veillonella*, also in the Neisseriaceae. These forms are harmless parasites of man.



Surface colonies of *Neisseria gonorrhoeae* and *N. catarrhalis* on chocolate agar, 48 hours. (From American Public Health Association, *Diagnostic Procedures and Reagents*, 3d ed., 1950)

Microscopically, in stained films from inflammatory exudates, *Neisseria* resemble paired coffee beans. They are characteristically seen engulfed within the cytoplasm of leucocytes. In pure culture, oval or spherical cocci are observed.

Neisseria are closely related antigenically and differentiation is important medically. This can best be done by observing differences in their biologic characteristics on culture media, although serologic reactions may help.

The ability of the various species to ferment simple sugars is a distinguishing characteristic. The Quellung reaction may be used to identify the meningococcus. Distinctive colonies are formed on suitable culture media by *N. catarrhalis* and *N. gonorrhea*. See GONORRHEA; GRAM'S STAIN; MENINGITIS; QUELLUNG REACTION. [C.M.C.]

Bibliography: W. Burrows et al., *Textbook of Bacteriology*, 17th ed., 1959.

Granite

A phaneritic (visibly crystalline) plutonic rock with granular texture composed largely of quartz and alkali feldspar with subordinate plagioclase and dark-colored (mafic) minerals (biotite and

hornblende). More generally the term includes most plutonic rocks rich in quartz and feldspar. In this sense it includes, in addition to true granite, such rocks as granodiorite and quartz monzonite. Commercially the term granite is still more inclusive and is used for any phaneritic rock rich in feldspar and with or without quartz and mafic minerals. Due to its strength, durability, and pleasing colors (gray, pink, and red) granite is an important building and ornamental stone. For a general discussion of textural, structural, and compositional characteristics *see* IGNEOUS ROCKS.

Mineralogy. True granite is distinguished from granodiorite by the preponderance of alkali feldspar over plagioclase. Otherwise the two rocks are similar. Alkali feldspar (microcline, orthoclase, usually perthitic) is generally pink, flesh, or buff and normally occurs as anhedral (without crystal form) or subhedral (with partial crystal outline) grains. In alkali granites it is more soda-rich than in normal (calc-alkali) granites. The common plagioclase of calc-alkali granites is oligoclase whereas alkali granites may carry discrete crystals of albite In granodiorite the plagioclase is commonly as calcic as sodic andesine. It is usually subhedral and may show normal zoning (calcium-rich cores surrounded by successively more sodic shells).

Quartz appears as glassy clear to smoky gray grains or grain aggregates of anhedral form and is generally interstitial. Under the microscope it may show effects attributed to strain. With high magnification it may be seen to contain tiny gas-liquid bubbles or inclusions and hairlike needles of rutile.

Black, flaky biotite mica (microscopically brown) is the principal mafic mineral of calc-alkali granites. It may occur as the only mafic or with more or less amphibole. Biotite of alkali granites is iron-rich. Green hornblende as short subhedral prisms is the usual amphibole, but soda amphibole (hastingsite, arfvedsonite, or riebeckite) is typical of alkali granites. Pyroxene is rare, but diopsidic augite may form cores within hornblende in calc-alkali granites. Alkali granites may carry aegirine-augite or aegirite as interstitial material or as acicular crystals Diopside and augite are more common and abundant in granodiorite.

Muscovite mica may be a primary or secondary mineral. Primary muscovite does not occur with hornblende or pyroxene but is usually accompanied by biotite. It may be intergrown with biotite or occur as thick hexagonal plates.

The chief accessory minerals of granites are magnetite, ilmenite, apatite, zircon, and sphene as tiny grains or crystals Dustlike particles of hematite are commonly abundant in the potash feldspar and give that mineral its pink or salmon color. A few granites carry fayalite, an iron-rich olivine, and some contain tourmaline or fluorite which may have formed subsequent to consolidation of the rock.

Texture. Granites display a wide variety of textures and structures. Cuneiform intergrowths of quartz and alkali feldspar forming micrographic or granophyric texture are not uncommon and characterize the special varieties, graphic granite and granophyre, respectively. A fine-grained, sugary textured rock (aplite) is developed in dikes and locally in some larger granite bodies. Elsewhere the texture becomes extremely coarse (pegmatitic) and the rock passes into granite pegmatite. *See* APLITE; PEGMATITE.

Many granites show porphyritic texture in which large crystals (phenocrysts) of alkali feld-par are embedded in a finer-grained matrix. The phenocrysts are euhedral (with good crystal form) to anhedral and may have formed either early or late in the period of crystallization. In some cases the large crystals formed by metasomatic replacement of solid rock and are more properly called porphyroblasts. Quartz is less common as phenocrysts. Fine-grained granites with abundant phenocrysts are known as granite porphyry. *See* GRANITIZATION; PHENOCRYST; PORPHYRY; RAPAKIVI GRANITES.

Structure. Directive structures are common in many granites and granodiorites (gneissic granites) and may be best developed or confined to margins of the body. These take such forms as banding, schlieren (elongate tabular lenses), and parallel orientation of elongate or platy minerals and inclusions. These features (flow structures) may trend parallel to the margin of the granitic body or may conform closely with similar structures in the surrounding rocks. They may have formed by flowage of the crystallizing granite magma or they may represent relics of bedding or foliation in metasomatic or metamorphic granite. Granites crystallized at shallow depth and, therefore, under relatively low pressure may show miarolitic cavities containing beautifully crystallized minerals. Orbicular structure, although not common, may be well displayed by granites and granodiorites.

Joints or extensive, smooth fractures in parallel arrangement are characteristic structures of most granitic rocks. A special type of fracture known as sheeting divides the granite into huge slabs or sheets, resting one above the other. Normally these fractures form parallel to the earth's surface, presumably due to expansion and release of confining pressure as erosion strips away the thick overburden. Closely spaced joints or joints which do not intersect at right angles may make a granite unsuitable for quarrying.

Occurrence. Granitic rocks form bodies of various sizes ranging from small dikes and sills up to huge batholiths covering thousands of square miles and localized in the cores of fold-mountain ranges and the great Precambrian shield areas of the world. Among the smaller and intermediate sized bodies are the stocks, sheets, domes, and a variety of irregular shaped plutons. *See* PLUTON.

In addition to occurring independently, granite may be associated with great masses of granodiorite and tonalite; with gabbro and intermediate

rocks in smaller complex assemblages; with gabbro and related rocks in large, thick, sheetlike bodies; with subordinate amounts of quartz syenite and syenite; and with small masses of quartz syenite and alkali syenite.

Origin. Just how granite forms constitutes a major problem in geology. Three principal types of processes appear to be operative (magmatic, metamorphic, and metasomatic); these may act independently or in various combinations. Magmatic granite forms by slow crystallization of a deeply buried granitic melt (magma). Metamorphic recrystallization (reconstitution by heat, pressure, and volatiles) may transform volcanic or sedimentary rocks into granite. A wide variety of sedimentary or igneous rocks may be changed to granite, in essentially the solid state, by the introduction of certain elements such as alkalies and silica and the removal of others such as iron, magnesium, and calcium. This process of replacement or metasomatism is involved in the phenomenon of granitization. *See* MAGMA; METAMORPHIC ROCKS; METAMORPHISM; PETROGRAPHIC PROVINCE. [C.A.CA.]

Bibliography: M. Walton, Granite problems, *Science*, 131(3401):635–645, 1960.

Granitization

The process whereby various types of rocks may be converted, essentially in the solid state, to granite or closely related material (such as granodiorite). The process is generally a large-scale change and is considered to have operated in orogenic zones (zones of fold-mountains) essentially contemporaneously with metamorphism. Here stratified rocks, undergoing reconstitution and recrystallization, may suffer changes in bulk composition adequate to convert them to rocks of granitic character. Such stratified rocks have thus been granitized, and the granites formed thereby may be indistinguishable from those of magmatic origin. In fact, one type of granite may pass imperceptibly into the other.

Granitization on a small scale, more or less confined to contacts of magmatic bodies at high levels in the earth's crust, might more appropriately be referred to as contact metasomatism or feldspathization. *See* AUREOLE, CONTACT.

The change in bulk composition during granitization involves the introduction of certain constituents and removal of others. This substitution, in essentially solid rock and with no appreciable overall volume change, is known as metasomatism. Bodily movement of fluids (along fissures and openings) carrying requisite substances in solution may account for the compositional changes in the rocks. Where openings are too small to permit fluid flow, diffusion along intergrain surfaces may occur. For short distances diffusion through the crystal structure may be important. Extensive deformation accompanying metamorphism may maintain adequate channelways for the migrating material. *See* METASOMATISM.

Movement of fluids, perhaps derived from some underlying body of magma or from the deep parts of the earth, would permit principally a one-way transfer. Elements necessary for granitization (such as alkalies) could be brought up from below; the displaced elements (calcium, Ca; iron, Fe; magnesium, Mg) could be carried to higher levels and, perhaps, the earth's surface.

Diffusion along intergranular surfaces or throughout a stationary pore fluid or diffusion through the crystal structure would permit multidirectional transfer. Each chemical component would move more or less independently, prompted by pressure and temperature gradients and by gravity. Consequently, certain components could move upward and others downward; but direction and rate of movement for a specific component could change with changes in physical conditions.

Some geologists believe that, as granitization proceeds, elements driven out may concentrate at the diffusion limit to form a zone or "front" with distinctive minerals. In some instances a basic front (rich in Ca, Fe, and Mg) appears to have formed in advance of a wave of granitization. According to the concept of two-way migration during granitization, it seems likely that basic material may move downward as alkalies and silica may move upward

Transformation may not proceed uniformly. It may take place more readily or completely along certain layers or beds so that long fingerlike processes of metamorphic rock are left extending deeply into metasomatic granite. Partially replaced layers may appear as narrow slivers and lenses enclosed by granite. These might resemble xenoliths in magmatic rocks drawn out and oriented by flowage. Evidence for metasomatic origin of granite is, perhaps, most compelling where bedding of the metamorphic rocks can be traced, though somewhat vaguely, well into the granite body. *See* XENOLITH.

The gradual transition from metamorphosed sediments to metasomatic granite may be studied by chemical and petrological methods. Mica schist or gneiss, derived from shaly sediments, may gradually gain in plagioclase and potash feldspar largely at the expense of muscovite mica and mafic (iron-magnesium) minerals. There is generally an increase in grain size and commonly a marked development of large feldspar crystals (porphyroblasts). This textural change may be accompanied by partial or complete obliteration of the original schistose or gneissose structure.

Large feldspar crystals in metamorphic inclusions in granite must be of replacement origin. They commonly resemble in detail the large crystals within the granite which also appear to have formed late. These large feldspars suggest, but do not prove, that the granite also is of metamorphic origin.

placement along certain beds or layers in metamorphic rocks could give rise to veined gneiss or banded varieties of migmatite. Replacement along diversely oriented fractures in a rock would at a certain stage create a migmatite composed of remnant blocks enclosed in granite matrix.

There is a mergence of metamorphic and igneous action where pods and streaks of magmalike fluid, rich in alkalies and silica (granitic), form in place by selective fusion or solution. Upon solidification of this interstitial material, another type of migmatite is developed. If before solidification, however, the rocks become compressed, the melt may be squeezed out to form large independent bodies of magma. It is possible that the whole mass may become mobilized and take on the aspect of a magma. The liquid phase serves to lubricate contacts between the more or less plastic, solid portions. Such mobilized mixtures are sometimes called migmas. Thus, migmatization (the formation of migma and migmatite), granitization, metamorphism, and igneous action may be closely associated phenomena *See* MAGMA; METAMORPHIC ROCKS; METAMORPHISM; MIGMATITE. [C.A.CA.]

Bibliography: H. H. Read, *The Granite Controversy*, 1957.

Granodiorite

A phaneritic (visibly crystalline) plutonic rock composed chiefly of sodic plagioclase (oligoclase or andesine), alkali feldspar (microcline or orthoclase, usually perthitic), quartz, and subordinate dark-colored (mafic) minerals (biotite, amphibole, or pyroxene). Granodiorite is intermediate between granite and quartz diorite (tonalite). Alkali feldspar is dominant over plagioclase in granite but is subordinate to plagioclase in granodiorite. Quartz diorite carries little or no alkali feldspar For convenience granite and granodiorite are commonly grouped and referred to as granite. *See* GRANITE; IGNEOUS ROCKS. [C.A.CA.]

Granulite

In petrology the term granulite has been used with different meanings: (1) denoting simply medium to fine-grained gneisses of more or less granitic composition; and (2) denoting banded, fine-grained metamorphic rocks having distinctive fabric (granulite texture, feldspar and particularly quartz deformed to thin plates) and mainly consisting of quartz and feldspar without mica but with a fair number of minute, rounded, pale-brown garnets. The second definition is now usually accepted for granulite.

Sometimes the first definition of the word is still applied. For example, a large part of the Highlands of Scotland consists of so-called granulites (Moine granulites) which are simply gneisses derived from arkose, grits, and sandstone.

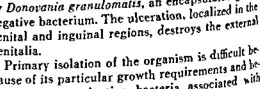
Typical granulites are described from Saxony and Finland (Lappland). They are banded with alternating layers of light granitic bands (the Weiss-stein of Saxony) and dark pyroxene-bearing layers or lenses, which in Lappland shows a noritic composition, but in Saxony exhibits great compositional variations indicating a sedimentary (calcareous, marly) parentage.

The metamorphism of these rocks has taken place at rather high temperature and pressure; thus the use of the term granulite facies has been proposed to designate the facies showing the great- [illegible] (high-grade [illegible] ks. [illegible] ly debated. The Saxon and Finnish rocks particularly have been studied in great detail (very similar rocks occur in Austria in the Wald-Viertel and in India). Structural relics from granites and quartz porphyries have been found, and it is now possible to assert (as maintained by E. Lehmann in 1884) that the granulites represent metamorphic—and to a large extent dynamo-metamorphic—products of a variety of older rocks. [T.F.W.B.]

Granuloma inguinale

An infectious, chronic, creeping ulceration caused by *Donovania granulomatis*, an encapsulated gram-negative bacterium. The ulceration, localized in the genital and inguinal regions, destroys the external genitalia.

Primary isolation of the organism is difficult because of its particular growth requirements and because of contaminating bacteria associated with ulcers. Diagnosis is best made by staining scrapings from the lesions with Wright's blood stain and examining for the typical intramonocytic organisms.



Mononuclear cell showing presence of Donovan bodies.

The microorganisms when seen in the cytoplasm of mononuclear cells in stained tissue smears are called Donovan bodies and when cultured on lifeless media are named *Donovania granulomatis*.

Transmission of the disease is undetermined. The infection is encountered chiefly in Negroes and occurs mostly in areas where the climate is warm and humid for several months of the year. Antibiotics have largely replaced antimony compounds for the

treatment of granuloma inguinale. Terramycin is probably the drug of choice. *See* BACTERIOLOGY, MEDICAL; BRUCELLACEAE. [R.B.D.; R.B.C.]

Grape culture

That division of horticulture concerned with grape growing. The broader term, viticulture, includes studies of grape varieties; methods of culture such as trellising, pruning, and training; and insect and disease control.

The genus *Vitis* is comprised mostly of plants which climb by means of the coiling of cylindrical-tapering tendrils. *See* LEAF (BOTANY). The flowers are polygamodioecious—some plants have perfect flowers, others have only staminate flowers with at most rudimentary pistils. *See* FLOWER (BOTANY). The 5 flower petals are greenish, narrow, and cohere at the top. The fruit is a pulpy berry. *See* FRUIT (BOTANY). In the Old World species, *Vitis vinifera*, probably a native of Western Asia, the skin and pulp of the mature berry cohere; in many North American species, the skin of the mature berries separates freely from the pulp.

Fig. 1. Foliage, tendrils, fruit, and stem characteristics of the Concord grape, *Vitis labruscana*. (From *E. L. Palmer, Fieldbook of Natural History*, McGraw-Hill, 1949)

Kinds and cultivation. Of the cultivated grapes of the world, more than 90% are clonal varieties of the species *V. vinifera*. *See* STEM (BOTANY). In the United States *V. vinifera* is grown on the West Coast. Muscat or Alexandria, Tokay, Emperor, Thompson Seedless or Sultanina, Cabernet, Pinot, and Riesling are among the leading varieties. Most of the grapes cultivated east of the Rocky Mountains have been derived from American wild vines such as *V. labrusca* and *V. aestivalis*, or from crosses between them and *V. vinifera*. The principal commercial varieties are Concord, Catawba, Isabella, Delaware, and Niagara. In the Southeastern and Gulf States varieties of *V. rotundifolia*, the muscadine grapes (scuppernong) are grown.

Many *vinifera* varieties are grown with a trunk 1–3 ft high bearing at the top a ring of short branches which are pruned to produce short spurs each having only a few buds. These trunks are staked in early years but become more woody and self-supporting when older. Other *vinifera* varieties and the American-type grapes are pruned to produce canes bearing 8–15 buds and are supported on a vertical trellis. The muscadine grapes are often grown on overhead arbors.

Production and value. World grape production probably exceeds that of any other fruit. The largest producers are Italy, France, Algeria, the United States, and Spain. These countries produce about 65% of the world total.

The major area of production in the United States is the West Coast. California, the center of production, produces about 80% of the total crop. In the East, the Chautauqua-Erie grape belt extends along the southeast shore of Lake Erie. Another area is found about the Finger Lakes of central New York.

The average annual production in the United States from 1948 to 1958 was second in value only to that of oranges (including tangerines), with apples a close third (*see* APPLE; ORANGE). The value of the crop in the United States has averaged about $130,000,000 yearly.

Uses. Of the world grape production, 80% is crushed for wine and 7% is dried as raisins. In the United States, about 40% is crushed and used for wine, brandy, or fresh juice, a part of which is processed into jams, preserves, or jellies; about 35% is dried as raisins; and about 25% is canned for use in fruit cocktail, frozen as a concentrated juice, or consumed as fresh fruit. *See* AGRICULTURAL SCIENCE (PLANT). [J.EI.]

Grape diseases. Diseases of European bunch grapes, *V. vinifera;* American bunch grapes, *V. labrusca*; and muscadine grapes, *V. rotundifolia*, vary in severity with the varieties planted and the climatic conditions. There are few regions in which grapes can be grown without disease control, and frequently diseases are limiting factors in production. A few fungus diseases are of major importance.

Black rot, *Guignardia bidwellii*, is the most serious grape disease throughout the world and is the only major disease that attacks muscadines. On muscadines only black scabs and cankers develop. On other grapes, however, symptoms are reddish-brown, circular spots on leaves, black cankers on stems, and a rot that reduces the berries to shriveled black mummies.

Downy mildew, *Plasmopara viticola*, is serious where the weather is warm and humid. Irregular, yellowed areas covered on the lower surface with a cottony white fungus growth appear on the leaves (*see* FUNGI). Young shoots are shortened, distorted, and covered with the downy growth, as are the young berry clusters. Older grapes turn brown, shrivel slightly, and drop.

Fig 2 Black rot leaf spot and tip blight of muscadine grape.

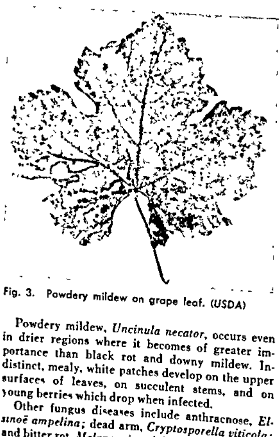


Fig. 3. Powdery mildew on grape leaf. (USDA)

Powdery mildew, *Uncinula necator*, occurs even in drier regions where it becomes of greater importance than black rot and downy mildew. Indistinct, mealy, white patches develop on the upper surfaces of leaves, on succulent stems, and on young berries which drop when infected.

Other fungus diseases include anthracnose, *Elsinoë ampelina*; dead arm, *Cryptosporella viticola*; and bitter rot, *Melanconium fuligineum*.

Bacteria. Blight of the shoots, *Erwinia vitivora*, and crown gall, *Agrobacterium tumefaciens*, cause minor damage.

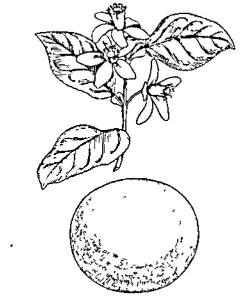
Viruses. Pierce's disease, caused by the alfalfa dwarf virus, damages vineyards in California and may be responsible for bunch grape failures in the southeastern United States. Leaves scald at the margins and the vines gradually decline and die. Other viruses produce mosaics, leaf-roll, witches' broom, and various growth distortions. (*See* PLANT VIRUS.)

Control of grape diseases. Cultural practices helpful in decreasing disease severity include selection of disease-free stock, planting on favorable sites with good air drainage, trellising, pruning, and cultivating to promote air circulation, and vineyard sanitation through removing diseased canes and litter. When fungus diseases threaten the crop, chief reliance must be placed on spraying. Powdered sulfur controls powdery mildew, but at high temperatures it injures the vines. Bordeaux mixture is the most generally effective fungicide for all diseases. Unfortunately, it also causes injury to the vines. Therefore, the organic fungicides, ferbam and zineb, are preferred for control of black rot but are less effective against mildews. *See* FUNGISTAT AND FUNGICIDE; INSECTICIDE; PLANT DISEASE CONTROL. [ESL]

Bibliography: *See* PLANT DISEASE.

Grapefruit

A species of citrus, *Citrus paradisi*, including both seedy and seedless varieties of fruits with pink, white, or red flesh. The trees are vigorous and evergreen with well-rounded tops (*see* EVERGREEN PLANTS). The grapefruit is the most recently developed of the citrus fruits, being first described in 1750. It is thought to have originated from another species, *C. grandis*, which includes the pummelos and shaddocks.



Foliage, flowers, and fruit of grapefruit, *Citrus paradisi*. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Of the world supply of grapefruit, about 90% is produced in the United States. Florida is the leading producer-state, with 116,000 acres in 1957, followed by Texas, California, and Arizona. From 1948 to 1958 the average yearly value at the packing house of all grapefruit in the United States was $43,469,000.

Approximately one-half of the grapefruit produced in the United States is marketed as fresh fruit, the rest is processed as canned juice and segments. *See* FRUIT (BOTANY); FRUIT (TREE); FRUIT (TREE) DISEASES. [F.E.C.]

Graph theory

A branch of mathematics which belongs in part to topology and more closely in subject matter and methods of proof to combinatorial analysis. It also occurs, sometimes under other names, in electrical network theory, organic chemistry, theoretical physics and statistical mechanics, and the group dynamics aspects of social psychology. J. J. Sylvester has described the combinatorial nature of graphs by stating that the theory of ramification is one of pure colligation because it takes no account of magnitude or position. Geometric lines are used, but have no more real bearing on the theory than those employed in genealogical tables have in explaining the laws of procreation.

Origin of graph theory. The beginnings of topology and simultaneously of graph theory, date from the celebrated Königsberg bridge problem. The problem was to find a way of walking across each

in which the points a, b, c, d represent land areas and the lines stand for bridges. He then proved that in any such problem it is possible to start at any point, cross each bridge exactly once, and return to the starting point if and only if every point of the corresponding graph is even (lies on an even number of lines). Since every point of G is odd, it follows that the Königsberg bridge problem is unsolvable. Graphs in which such traversability problems are solvable are now known as Euler graphs. If exactly two points of a connected graph are odd, then there is still a sequence which includes all lines, starting from one of these odd points and ending at the other.

The so-called Euler characteristic arose in the following context. Consider any polyhedron drawn on a sphere and let V, E, and F be the number of vertices, edges, and faces, respectively; for example, for a cube, $V = 8$, $E = 12$, and $F = 6$. Euler proved that $V - E + F = 2$ This is equivalent to a theorem on graphs which has been generalized in combinatorial topology to the Euler-Poincaré characteristic for simplicial complexes. *See* TOPOLOGY.

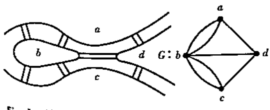


Fig. 1. The seven bridges of Königsberg

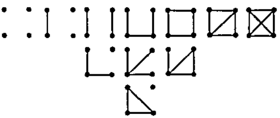


Fig. 2. Diagrams of all 4-point graphs.

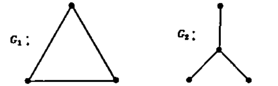


Fig. 3. Two graphs with three mutually adjacent lines.

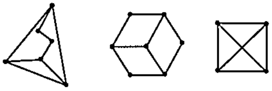


Fig. 4. Homeomorphic but nonisomorphic graphs.

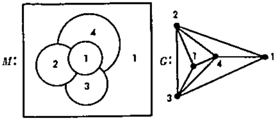


Fig. 5. A map requiring four colors, indicated by numbers, and its planar graph.

Definitions. A graph consists of a finite collection of points together with a set of lines joining pairs of distinct points. Two points of a graph are adjacent if they are joined by a line; each point is incident to the line. Two lines are adjacent if they have a common point. A path is a sequence of successively adjacent lines in which all points are distinct; a cycle is obtained on joining the end points of a path. A graph is connected if there is a path between any two points. A tree is a connected graph with no cycles. In a complete graph, any two points are adjacent. Two graphs are isomorphic if there is a one-to-one correspondence between their sets of points which preserves the adjacency relation.

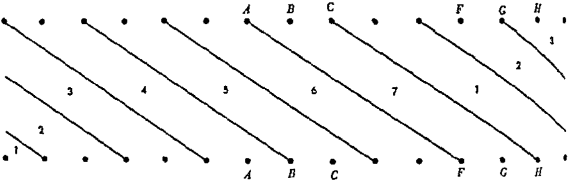


Fig. 6. A map on a torus (opposite sides of the rectangle are identified) requiring seven colors.

Hassler Whitney has shown that if there is a one-to-one correspondence between their sets of lines which preserves adjacency, then two graphs are isomorphic unless they are the two graphs G_1 and G_2 consisting of three mutually adjacent lines, shown in Fig. 3.

An automorphism of a graph is an isomorphism with itself. The degree of a point is the number of lines incident to it. Two graphs are homeomorphic if, on suppressing all points of degree 2, the resulting graphs are isomorphic.

The four-color conjecture. Given any map in the plane, it is possible to color the regions with four colors so that any two regions with a common boundary arc have different colors. This is a famous conjecture which to date has been proved for all maps with fewer than 36 regions.

Surprisingly, on surfaces more complicated than the sphere, such as a doughnut or a pretzel, the coloring problem is completely solved. For exam- [illegible]

AB, and 5 along BC.

In a sphere or a plane, only the 5-color theorem has been proved. A graph is n-chromatic if it requires n different colors in order that no two adjacent points have the same color; for example, the complete n point graph, K_n. G. A. Dirac has shown that every 4-chromatic graph contains a subgraph homeomorphic to K_4. The assertion for $n = 5$ implies the 4-color conjecture and is therefore hopeless at present. When a graph is drawn on a surface, its points and lines are called vertices and arcs. Planar graphs are those which can be drawn in a plane with none of its arcs intersecting each other.

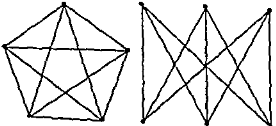
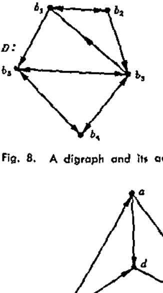


Fig. 7. Prototypes of all nonplanar graphs.

Kuratowski's theorem states that a graph is planar if and only if it contains no subgraph homeomorphic to either of the two graphs shown in Fig. 7.

The second graph in Fig. 7 is the subject of a well-known impossible puzzle, namely, to join in the plane each of three houses with each of three utilities with none of the joining arcs intersecting each other. However, it is easy to draw this graph on a torus.



$$M = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Fig. 8. A digraph and its adjacency matrix.

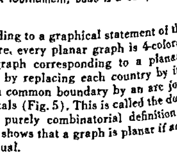


Fig. 9. A tournament; *badc* is a complete path.

According to a graphical statement of the 4-color conjecture, every planar graph is 4-colorable. The planar graph corresponding to a planar map is obtained by replacing each country by its capital and each common boundary by an arc joining the two capitals (Fig. 5). This is called the *dual graph*. Using a purely combinatorial definition of dual, Whitney shows that a graph is planar if and only if it has a dual.

A cubic graph has only points of degree 3. Most of the attempts at solving the 4-color conjecture lead only to equivalent restatements, often in different form, for example, Tait's conjecture: every planar cubic graph can have its lines colored with 3 colors.

In a directed graph, or digraph, each line $\overrightarrow{ab}$ is directed from point a to another point b. The adjacency matrix $M = (m_{ij})$ of a digraph D with points $b_1, b_2, \ldots, b_n$ has $m_{ij} = 1$ if the line $\overrightarrow{b_ib_j}$ occurs in D; otherwise $m_{ij} = 0$.

An oriented graph is obtained from an ordinary graph by assigning a unique direction to each line. A complete oriented graph or, more briefly, a tournament, is obtained by orienting a complete graph. A complete (or hamiltonian) path in a graph contains all its points. According to Redei's theorem, every tournament has a complete directed path. Further, the number of complete paths in any tournament is odd.

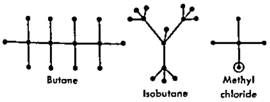


Fig. 10. The two isomers of C_4H_{10}, and a monosubstituted hydrocarbon.

An automorphism of a graph G is an isomorphism of G with itself. The group of a graph is the set of all its automorphisms; for example, the group of K_n is S_n, the symmetric group of degree n. The previously open problem of deciding whether a given abstract finite group is isomorphic to the group of some graph was settled in the affirmative by R. Frucht.

Applications. The number of isomers of saturated hydrocarbons (C_nH_{2n+2}) was expressed by Arthur Cayley as the graphical problem of enumerating the class of "trees" in which each point has degree 1 or 4 (for a hydrogen or carbon atom respectively).

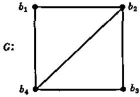
An elegant and general solution was first obtained by G. Pólya, who developed a powerful enumeration technique which also serves to handle many related problems. Extremely effective use of Pólya's theorem has been made by the theoretical physicists R. W. Ford and G. E. Uhlenbeck in solving several enumeration problems for graphs which arise in a study of statistical mechanics. *See* ISOMERISM, MOLECULAR.

According to Menger's theorem, if G is a connected graph and A and B are disjoint sets of points of G, then the minimum number of points whose deletion separates A and B is equal to the maximum number of disjoint paths between A and B. This elegant theorem in pure mathematics is closely related to applications to linear programming of the "assignment," "transportation," and "traveling salesman" problems. *See* LINEAR PROGRAMMING.

Kirchhoff's laws are fundamental in electrical network theory. While developing his equations for the current in each branch of an electric network in terms of input voltages, G. R. Kirchhoff was led to the following classical result. A spanning subgraph of G has the same points as G. According to Kirchhoff's matrix tree theorem, if G is a connected graph with points $b_1, \ldots, b_n$, and if $A = (a_{ij})$ is the matrix in which $a_{ij} = -1$ whenever b_i and b_j are adjacent and 0 otherwise, while a_{ii} is the degree of b_i, then all cofactors of A are equal to the number of spanning subtrees of G.

Signed graphs are obtained from ordinary graphs by taking some of the lines as positive and the remaining lines as negative. The concept of "balance" in signed graphs and its psychological relevance are explored by D. Cartwright and F. Harary.

Group dynamics is a branch of social psychology which studies problems involving the structure of a group. If points represent people and lines their interrelationships, then a graph may be used to depict the structure of a social group. The relations studied in psychological problems include liking, communication, power, and their negatives. The types of constructs in social groups for which a graph theoretic approach proves fruitful are illustrated by liaison persons, cliques, and strengthening and weakening members of an organization.



$$A = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}$$

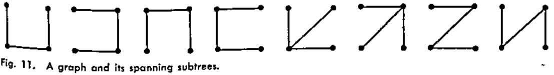


Fig. 11. A graph and its spanning subtrees.

See Boolean algebra; Conformal mapping; Lattice (mathematics); Polytopes, regular; Ring theory; Set theory. [F.H.]

Bibliography: D. Cartwright and F. Harary, Structural balance, *Psychol. Rev.* 63(5):277–293, 1956; H. S. M. Coxeter, Map-coloring problems, *Scripta Mathematica*, 23:11–25, 1957; G. A. Dirac, A property of 4-chromatic graphs, *J. London Math. Soc.*, 27:85–92, 1952; L. Euler, The Königsberg bridge problem, *Sci. American*, 189(1):66–70, 1953; G. W. Ford and G. E. Uhlenbeck, Combinatorial problems in the theory of graphs, I–IV, *Proc. Nat. Acad. Sci.*, 42:122–128, 203–208, 529–535, 1956, 43:163–167, 1957; R. Frucht, Graphs of degree three with a given abstract group, *Canad J. Math.*, 1(4):365–378, 1949; F. Harary, Graph theoretic methods in the management sciences, *Management Science*, 5:387–403, 1959; F. Harary, Graph theory and electric networks, *Inst. Radio Engr. Trans. on Circuit Theory*, CT-6:95–109, 1959; F. Harary and R. Z. Norman, *Graph Theory as a Mathematical Model in Social Science*, 1953.

Graphic methods

Pencil and paper methods that employ the geometry of a plane to express mathematical relationships and to carry out mathematical operations in analog form. Some of the specific topics that arise under this broad definition are the plotting of experimental data, curve fitting and development of empirical formulas, graphical interpolation and extrapolation, nomographs and alignment charts, graphical differentiation and integration, and graphical statics.

Since graphical methods are often rapid and easily understood, they find widespread application in engineering. Thus, for example, the stresses in the individual members of a bridge can be determined by the methods of graphical statics which use arrows or vectors to represent the direction and magnitudes of the forces. The latter are specified by the lengths of the arrows. Equilibrium of the structure requires that all the vectors representing the forces acting on any part of the structure should, when placed end to end, form a closed polygon. Using similar graphical formulation for all the equilibrium conditions, one may by graphical construction determine the forces or stresses in the individual members in terms of the loads applied to the bridge.

Graphical techniques are also useful in statistical studies to determine the correlation between two quantities, to obtain a frequency distribution from experimental data, to smooth data having an appreciable statistical fluctuation, and to represent the conclusions of a statistical study. *See* Coordinate systems, graphical; Curve fitting; Extrapolation; Interpolation; Nomograph; Numerical analysis; Statics; Statistics. [K.S.K.]

Bibliography: D. S. Davis, *Nomography and Empirical Equations*, 1955; F. A. Willers, *Practical Analysis*, 1949.

Graphite

A mineral consisting essentially of carbon but often impure with clay, iron oxide, or other minerals. Two forms of elemental carbon occur as minerals, graphite and diamond. Graphite crystallizes in the hexagonal system. Crystals are rare but when observed are small six-sided tablets with 60° triangular markings on the broad faces. Graphite commonly occurs in foliated masses but may occur as radiated, granular, earthy, or columnar aggregates.

Properties. Graphite is one of the softest minerals, with hardness 1–2 on the Mohs scale. It soils the fingers when handled, leaves a black mark on paper, and has a greasy feel. Its specific gravity is 2.09–2.23. The color is black to steel-gray, and the thinnest flakes are a deep blue in strong transmitted light. There is one perfect cleavage allowing the mineral to be split into thin, flexible, but nonelastic folia. Graphite conducts electricity, a property which permits it to be distinguished from amorphous carbon.

Physically, graphite strongly resembles molybdenite (molybdenum sulfide) with which it has been confused for centuries. Molybdenite crystallizes in the hexagonal system, is platy, and soft enough to leave a mark on paper. The mark left by molybdenite on glazed paper is greenish, whereas the mark left by graphite is black. Graphite has also been confused with lead sulfide, galena; hence the trade names plumbago and black lead both refer to graphite.

Occurrence. Graphite occurs chiefly in those rocks that have undergone intense metamorphism, and thus is found in the older gneisses, schists, and crystalline limestones. Carbonaceous material in original sandstone, shale, and limestone recrystallizes in the form of graphite. Such "crystalline" graphite has been mined in the Adirondack region of New York. Bedded deposits in Korea and Rhode Island have been converted from coal to graphite during metamorphism. Graphite may also be the product of contact metamorphism where igneous rocks intrude carbonaceous sedimentary rocks. In Sonora, Mexico, granite has intruded several coal beds, converting them to graphite. In Ceylon, graphite occurs in veins 6–10 in. wide in which the graphite has a fibrous structure with the fibers

industry
ensable

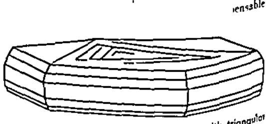


Fig. 1. Hexagonal crystal of graphite with triangular markings on face. (*From C. Palache, H. Berman, and C. Frondel, Dana's System of Mineralogy, vol 1, 7th ed., Wiley, 1944*)

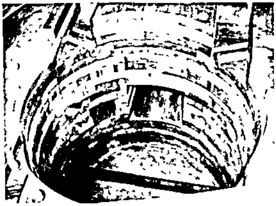


Fig. 2. Graphite reflector for a test reactor.

for certain purposes. The principal producing countries are Mexico, Korea, Austria, Madagascar, United States, and U.S.S.R. [C.S.HU.]

Synthetic graphite. Graphite has a highly developed crystalline structure, and its softness, high thermal and electrical conductivity, and self-lubricating qualities differentiate it from other forms of carbon.

Carbon in graphitic form has both metallic and nonmetallic properties. Commercially produced synthetic graphite is a mixture of crystalline graphite and cross-linking intercrystalline carbon. Its physical properties are the result of contributions from both sources. Thus, among engineering materials, synthetic graphite is unusual because a wide variation in measurable properties can occur without significant change in chemical composition.

At room temperature the thermal conductivity of synthetic graphite is comparable to that of aluminum or brass. An unusual property of graphite is its *increased strength at high temperature.* The crushing strength is about 20% higher at 1600°C and the tensile strength is 50–100% higher at 2500°C than at room temperature.

Graphite is resistant to thermal shock because of its high thermal conductivity and low elastic modulus. It is one of the most inert materials with respect to chemical reaction with other elements and compounds. It is subject only to oxidation, reaction with and solution in some metals, and formation of lamellar compounds with certain alkali metals and metal halides.

Uses. Graphite has many uses in the electrical, chemical, metallurgical, nuclear, and rocket fields: electrodes in electric furnaces producing carbon steel, alloy steel, and ferroalloys; anodes for electrolytic production of chlorine, caustic and chlorates, magnesium, and sodium; motor and generator brushes; sleeve-type bearings; seal rings; electronic tube anodes and grids; nuclear reactor moderators, reflectors, and thermal columns; rocket motor nozzles; missile rudder vanes; metallurgical molds and crucibles; linings for chemical reaction vessels; and in the resin-impregnated impervious form, for heat exchangers, pumps, piping, valves, and other process equipment.

Preparation. Synthetic graphite can be made from almost any organic material that leaves a high carbon residue on heating to 2500–3200°C. In commercial operations, raw materials are carefully selected because not all substances with high carbon content undergo a suitably complete transformation to graphite at these temperatures. Petroleum coke is raw material for the most commonly used production process. After calcining and sizing, the coke is mixed with coal-tar pitch, heated to about 165°C and formed by extrusion or molding to "green" shapes. Baking to 750–1400°C in gas- or oil-fired kilns follows the forming operation. Graphite is produced by heating the baked shapes to 2600–3000°C by passing electricity amounting to 1.6–3.0 kw per pound of graphite through the bed of a furnace made of the shapes laid in granular coke (Fig. 3). The whole bed is covered by an insulating blanket of silicon carbide, coke, and sand. Higher-density synthetic graphite can be obtained by impregnating the baked carbon with pitch prior to graphitization. Graphite with total ash content less than 20 parts per million is needed

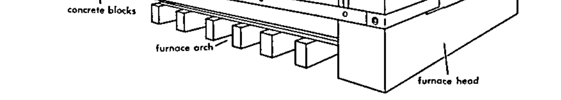


Fig 3. Graphitizing furnace.

for a number of nuclear and electrolytic uses, and is obtained by heating the graphite shapes electrically to about 2500°C while bathing them in a purifying gas. See CARBON; COKING (PETROLEUM REFINING); DIAMOND; ELECTROCHEMICAL PROCESS. [W.M.C.]

Bibliography: L. M. Currie, et al., The production and properties of graphite for reactors, *Proc. Int. Conf. Peaceful Uses of Atomic Energy*, 8:451–473, 1956; J. E. Hove, Graphite as a high temperature material, *Trans. Met. Soc. AIME*, 212(1): 7–13, 1958; R. E. Kirk and D. F. Othmer (eds.), *Encyclopedia of Chemical Technology*, vol. 3, 1949; National Carbon Co., *The Industrial Graphite Handbook*, 1959.

Graptolithina

A class of extinct colonial animals believed to be related to the class Pterobranchia of the subphylum Hemichordata. They are found in marine Paleozoic rocks of all continents but are most abundant in Ordovician and Silurian rocks where species are found to be characteristic of different series of strata (zones). These zones may be used for world-wide stratigraphic correlation. See HEMICHORDATA.

Graptolites (also known as Graptozoa) have been classed most often as Coelenterata, but the stolon system and structure of the skeleton point strongly to affinities with the pterobranchs such as *Rhabdopleura*.

The class Graptolithina is divided into several orders, of which the Dendroidea and Graptoloidea are the most important. The other orders and some doubtfully associated groups are known from relatively few occurrences.

Dendroidea. The dendroid graptolites are commonly sessile, [illegible] form and branching of the colony. Individual genera are widespread geographically, but the species appear to be quite local and are of little or no use for stratigraphic correlation. Detailed studies have been made of some dendroids etched out from chert and limestone. The organic skeleton is probably of some form of chitin. Two types of polyp are present in the colony and the individual cups, or thecae, are connected by a system of fine tubes (stolons) inside the colony (Fig. 1). The thecal wall consists of two layers, an inner fusellar layer showing growth bands and a much thicker cortical layer outside. The large autothecae and the smaller bithecae generally open together to the outside and probably represent individuals of different sex. In all graptolites the colony is produced by asexual budding from a single polyp (stolotheca), presumably the result of sexual reproduction, housed in the initial cup (sicula) which is generally quite distinct from the later thecae (Fig. 2).

Dendroid colonies frequently attain considerable size. Fan-shaped structures 10 cm across are not uncommon, but the individual thecae are small. In some genera the thecae are long but slender tubes which overlap their neighbors for the greater part of their length. In others, the thecae are short, wide tubes with little overlap.

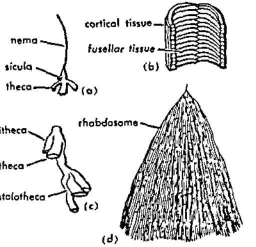


Fig. 1. Diagram showing structure and terminology of a dendroid graptolite. (a) Beginning of colony (b) Thecal wall. (c) Portion of a branch. (d) Whole colony.

Dendroidea are found from the Upper Cambrian to the Mississippian but most specimens are poorly preserved carbonaceous films from which only the general shape of the colony can be determined. The detailed structure is known for relatively few species and so cannot yet be used to provide more exact diagnostic characters which might lead to some stratigraphic use.

Graptoloidea. The order Graptoloidea is much more restricted in its stratigraphic range, being confined to the Ordovician and Silurian. It contains numerous genera and species of stratigraphic value. Although the colonies, or rhabdosomes, are occasionally associated with a shelly fauna in limestones and sandstones, most graptoloids have been described from black shales in which they form almost the whole fauna. The mode of formation of

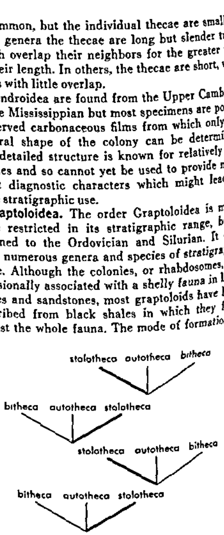


Fig. 2. Diagram indicating development of dendroid graptolite in alternating triads.

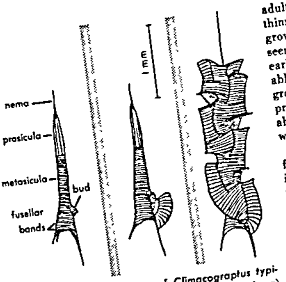


Fig. 3. Three growth stages of *Climacograptus typicalis* Hall, Ordovician, Ohio. (After O. M. B. Bulman)

these graptolitic shales is still under discussion, but it is agreed that the graptolites themselves were planktonic, probably attached to floating seaweed. In only a few cases have structures been found which can be interpreted as floats.

The development of the rhabdosome is known for about 20 species. By breaking up graptolite-bearing limestones in acid, large numbers of growth stages have been obtained as well as fragments of adult colonies. The outer cortex is usually much thinner in graptoloids than in dendroids, and the growth bands of the fusellar layer can be easily seen. The sicula (initial zooid) is in two parts. The earlier prosicula has no growth bands and is probably a larval covering. The later metasicula has the growth bands of a simple theca. The prosicula is prolonged into a thread (nema) which was probably the means of attachment to the floating seaweed (Fig. 3).

The earliest graptoloids were many-branched forms, found only in the lowest Ordovician. There is a general tendency in later forms to reduction in the number of branches. The branches (stipes) in the early forms are generally pendant or horizontal from the sicula. Another trend of evolution leads to upward growth along the nema, giving biserial or quadriserial rhabdosomes (two or four rows of thecae). The upward growth causes changes in the development of the first few thecae as has been shown by the detailed study of well preserved specimens. These differences, while important in the phylogenetic classification of the graptolites, cannot be seen in most black shale specimens. Thus the usual classification and stratigraphic studies rely on the grosser characters of thecal shape, thecal number, and the type of rhabdosome.

The Ordovician graptoloids have been subdivided into a large number of genera, based mainly on the branching and form of rhabdosome as illustrated in Fig. 4. Most of the Silurian forms have been placed in the large genus *Monograptus* which is subdivided according to varying thecal shapes and rhabdosome coiling. In the Middle Silurian there was a reappearance of branched forms of the genus *Cyrtograptus*. [I.S.T.]

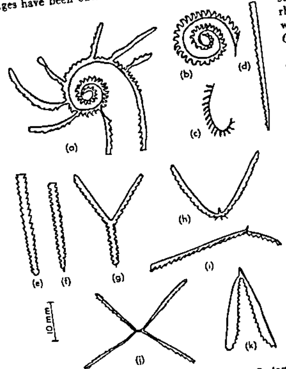


Fig 4. Shape of graptoloid colonies. (a) *Cyrtograptus*; (b, d) *Monograptus*; (c) *Rastrites*; (e) *Orthograptus*; (f) *Climacograptus*; (g) *Dicranograptus*; (h) *Dicellograptus*; (i, k) *Didymograptus*; (j) *Tetragraptus*.

Bibliography: O. M. B. Bulman, *Treatise on Invertebrate Paleontology*, pt. V, Geol. Soc. Am., 1955; R. R. Shrock and W. H. Twenhofel, *Principles of Invertebrate Paleontology*, 2d ed., 1953.

Grass crops

The primary livestock feed utilized as pasture and range forage, hay, and green feed or grass silage. Grasses in pasture, meadow, range, and turf occupy approximately 1,000,000,000 acres, or 75% of the agricultural land in the United States (Fig. 1). The annual value of grass crops is estimated from $5,000,000,000 to $10,000,000,000.

Structure. Grass stems have solid joints (nodes) and leaves arranged in two rows, with one leaf at each joint (Fig. 2). The leaves consist of the sheath, which fits around the stem like a split tube, and the blade, which is commonly long and narrow. Seed heads are made up of minute flowers on tiny branchlets, often several crowded together, but always two-ranked like the leaves. The flowers are generally wind pollinated. The seeds are enclosed between two bracts or glumes which remain on the seed when ripe.

Growth characteristics and distribution. Grass species may be annual or perennial. Some perennial species are perpetuated by creeping stems as wel

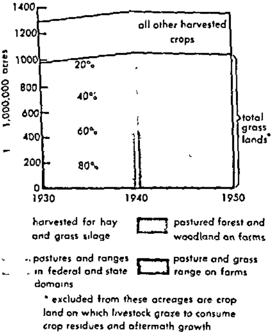


Fig. 1. Importance of grasslands in the United States. (From H. B Sprague, *The importance of grasslands, Agr. Food Chem., 2(21):1064, 1954*)

as by seed. Stems that creep underground are termed rhizomes, or rootstocks, whereas surface-creeping stems are called runners, or stolons. Species having creeping stems produce a well-knit sod These generally tolerate continuous close grazing and are utilized primarily for pastures. Species that do not creep vegetatively tend to grow erect. They make an open or loose sod, and are best suited for hay or silage. In subhumid to semiarid regions, range species are predominantly perennial and are perpetuated mostly by seed. In other areas both annuals and perennials commonly occur. All grasses have a fibrous type of root system that permeates the soil extensively and is effective in preventing soil erosion and restoring soil humus content. Grasses are world-wide in distribution; adapted species being available for nearly all conditions of soil and climate. See separate articles for grasses listed under their common names. [H.B.S.]

Diseases of grasses. Grasses are so commonplace and abundant that, until recently, little heed was paid to agents that killed or injured them However, intensified use on farms and turf areas has made the problem of controlling grass diseases one of considerable importance.

Kinds of pathogens. Grasses are attacked by several hundred kinds of pathogens. No one grass is

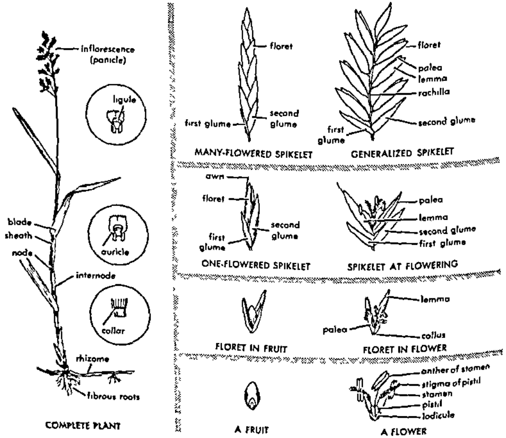


Fig 2. Parts of a typical grass plant. (From P. D. Strausbaugh and E. L. Core, *Flora of West Virginia, West Va. Univ. Bull., ser. 52, no. 12–2, p. 67, pt 1, 1952*)

infected by all, but many grasses are susceptible to
30 or more different disease-causing agents. Fungi
cause most of the diseases but bacteria, viruses,
and nematodes also attack grasses. The prevalent
fungi that attack the leaves of grasses cause the
most conspicuous symptoms. They reduce produc-
tivity and nutritive value of forage and may result
in depletion of stands. Some of the fungi that in-
cite rusts, smuts, and root rots of grasses also at-
tack cereals. Such fungi as those causing stem rust,
scolecotrichum brown stripe, and anthracnose of
grasses are adapted to a wide range of climatic
conditions and can infect numerous species of
grass. Such others as stripe rust, stem smuts, and
some leafspot fungi are more restricted in host
range and distribution.

Some of the viruses that attack grasses also in-
fect cereals and other crop plants. For example,
both the barley yellow-dwarf virus and the wheat-
reak mosaic virus attack numerous grasses. The
irus that causes dwarfing in alfalfa and Pierce's
lisease in grapevine also occurs in Bermuda grass.

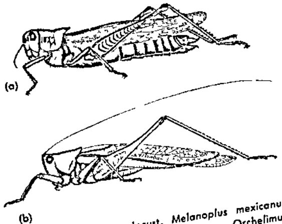
Control measures. Application of fungicides for
control of grass diseases is not considered practi-
cal except for some lawn and turf grass diseases,
and for certain fungi which attack grasses grown
for seed production. Consequently, control has been
obtained primarily by breeding and selection of
resistant varieties or through the use of improved
management practices. For example, Sudan grass
(*Sorghum vulgare* var. *sudanense*), orchard grass
(*Dactylis glomerata*), and Kentucky bluegrass
(*Poa pratensis*) can be damaged severely by sev-
eral foliar diseases. Varieties resistant to some of
these diseases have been developed through breed-
ing and selection. Postharvest burning of grass seed
fields in western United States is an example of
disease control by proper management practices
This technique has given practical control of such
diseases as blind seed (*Phialea temulenta*) of per-
ennial rye grass and leaf rust (*Puccinia poae-
sudeticae*) of Kentucky bluegrass. See NEMATODA;
PLANT DISEASE CONTROL; PLANT VIRUS. [K.W.K.]

Bibliography: *Grass*, USDA Yearbook Agr., 1948.

Grasshopper

Any of a large number of insects belonging to the
order Orthoptera, including all of the family Acri-
didae and many members of the family Tettigoni-
idae. Older classifications, however, frequently use
the family names Locustidae and Tettigidae for the
grasshoppers. The Acrididae have short antennae
and the ovipositor is short and inconspicuous; the
Tettigoniidae are distinguished by their antennae,
usually longer than the body, and the long, sword-
like ovipositor of the female.

Many of man's most serious enemies are in-
cluded in the group. The Biblical locust was a
member of the Acrididae, and is still an agricul-
tural problem of great importance over much of
Africa. In America there have been a number of
destructive outbreaks of grasshoppers, notably the



(a) Lesser migratory locust, *Melanoplus mexicanus*; length 1 in (b) Meadow grasshopper, *Orchelimum vulgare*; length to 2 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Rocky Mountain locust, the best-known of the
migratory species. The large, wingless Mormon
cricket, of the family Tettigoniidae, is the animal
famed for the destructive attack on the Mormon
crops in Utah in 1848, and is still a major pest in
the Great Basin. In the Middle West the differential
grasshopper annually causes considerable crop
loss and is fought with a variety of insecticides.
See INSECTICIDE.

Grasshoppers generally overwinter in the egg
stage. Eggs of Acrididae are usually deposited in
the ground; those of the Tettigoniidae are often
deposited on twigs, stems, or leaves, under bark,
or in the ground The young resemble adults ex-
cept that they lack wings. Their development is of
the type known as gradual metamorphosis; the
adult stage is slowly attained through a series of
molts. There is no resting stage.

Grasshoppers are chewing insects. They feed
almost indiscriminately on any available plant.
Aldrin and Dieldrin are the most effective of the
recent insecticides; Toxaphene and poison bait are
also frequently used in their control.

The term locust is commonly used in this country
to indicate the cicada, but the name locust is
properly limited to the grasshoppers. The lubber
is a large, heavy-bodied, flightless form of grass-
hopper. See INSECTA; KATYDID; ORTHOPTERA.
[J.D.B.]

Gravel

A gravel is a loose, or unconsolidated, deposit of
rounded pebbles, cobbles or boulders, whose size
is greater than 2 mm in diameter. Gravels may
consist only of pebbles, cobbles and boulders, with
large voids between the particles; these are the
openwork gravels. Commonly, however, gravels
have sand filling the interstices between the peb-
bles. The sand filling the pores is referred to as the
matrix and commonly makes up about 30% of the
rock. The size distribution of gravels has been
used as a clue to origin. Beach gravels normal
seem to have only one modal size (that size a

which most of the particles cluster) but river gravels have two modes, one in the gravel size and one in the sand sizes. The shapes, roundness, and surface textures of pebbles have also been used for guides to origin. Thus the degree of roundness and sphericity reflects the rigors of transportation and may be directly related to distance traveled. The shapes of pebbles may be inherited from fracture, joints, cleavage, and bedding patterns of the parent rock. Faceted pebbles called dreikanters have been produced by wind action. The shape of these pebbles is probably related to the pebble rock type and its bedding and jointing characteristics, both of which affect the original shape of the particle before transportation. *See* BRECCIA; CONGLOMERATE; SEDIMENTARY ROCKS. [R.S.]

Gravimetric analysis

That branch of quantitative chemical analysis in which a desired constituent is converted (usually by precipitation) to a pure compound or element of definite, known composition, and is weighed. In a few cases, a compound or element is formed which does not contain the constituent but bears a definite mathematical relationship to it. In either case, the amount of desired constituent can be determined from the weight and composition of the precipitate. The following are the essential steps in a conventional gravimetric analysis.

Dissolution. A sample is weighed and dissolved in a suitable solvent. *See* BALANCE (ANALYTICAL). Water and dilute mineral acids dissolve most inorganic substances, but occasionally concentrated acids or the specific effect of hydrofluoric acid are required. Some refractories require fusions to convert them to acid-soluble products. *See* SOLUBILIZING OF SAMPLES.

Precipitation. After removal of any interfering substances, the desired constituent is precipitated by the addition of the appropriate reactant to the properly prepared solution. Conditions are regulated so that coprecipitation of foreign substances is minimized. In most cases, favorable conditions are attained by precipitating the constituent from a well-stirred, highly diluted, hot solution by the drop-by-drop addition of a dilute reagent in only slight excess over the amount theoretically required. In order to obtain a more nearly pure product, it is usually desirable to dissolve the precipitate in a suitable solvent and to form it a second time. *See* PRECIPITATION (CHEMISTRY).

Digestion. In order to be readily filterable, a suspension of a precipitate is ordinarily allowed to stand at high temperature for the period of time necessary to permit amorphous particles to clot, or crystalline particles to increase in size. This digestion is usually carried out on an electric hot plate so regulated as to maintain a temperature just below the boiling point of the liquid. The addition of paper pulp is helpful in aiding subsequent filtration, but may be used only in cases where the precipitate is to be ignited at a temperature that causes the pulp to burn off.

Filtration. Filtration is accomplished by pouring the suspension of a precipitate through a suitable filtering medium. Whatever the medium, as much of the supernatant liquid as possible is decanted through the filter, and the transference of the precipitate is delayed as long as possible. The common filters are as follows:

Filter paper for gravimetric use is specially prepared and has undergone a treatment with hydrofluoric acid and with other acids so that on ignition it gives an ash of known, and usually negligible, weight. Papers of different degrees of porosity are available, and in a given filtration that grade is chosen which gives as rapid a filtration as possible and yet retains the precipitate completely. In general, gelatinous precipitates require a coarse-mesh paper; fine crystalline precipitates require a fine-mesh paper. Suction is almost never applied in paper filtration.

A Gooch crucible is a perforated-base porcelain crucible, and the filtering medium is a pad of asbestos produced by pouring into the crucible a suspension of asbestos fibers which are matted by applying suction. A perforated plate on top of the pad holds it in place. Fiber-glass disks may be used instead of asbestos. Suction is applied during filtration; the crucible is dried to constant weight and is weighed before and after the filtration. Gooch crucibles are usually used for precipitates that attain a definite composition at the moderate temperature of a drying oven.

A Munroe crucible differs from a Gooch crucible in that it is made of platinum and uses a platinum sponge (produced by igniting alcohol-moistened ammonium chloroplatinate) as the filtering medium. It retains even fine precipitates and can be heated to a high temperature. It is too expensive for routine use.

An Alundum crucible is made from aluminum oxide and is porous throughout. It therefore needs no additional filtering medium and can be heated to a high temperature.

A Selas crucible has glazed porcelain sides and an unglazed porcelain base which serves as the filtering medium.

A glass filtering crucible has a sintered glass base for the filtering medium and is favored in cases where the precipitate needs to be heated only to the moderate temperatures of a drying oven.

Washing. Precipitates are washed (usually with hot water) until essentially free from soluble foreign matter. Occasionally an aqueous solution of an electrolyte is used to prevent peptization of the precipitate, with resulting conversion to a colloidal solution. Ammonium nitrate is favored for this purpose since it is removed by volatilization when the precipitate is subsequently ignited. Washing is more efficient if the wash water is added in several portions with intermediate drainage.

Drying and ignition. Some substances can be dried to constant weight by heating to relatively low temperatures (110–275°C) in drying ovens. High-temperature ignition is usually carried out

on a precipitate that has been filtered on paper. First the paper is smoked off at a low temperature, and the residue is ignited either in an electric muffle furnace or over a free flame. In the latter case, one of the following types of burners is used:

A bunsen burner is a simple tube with an opening at the base to permit air to be mixed with the illuminating gas used.

A tirrill burner is a modification of the bunsen burner and allows greater flexibility in the adjustment of the air-gas mixture.

A meker burner is larger in diameter than the tirrill burner and has a grid at the top to give a broad, hot flame.

A blast lamp is a burner that supplies air (or oxygen) under pressure to the illuminating gas used.

The temperature delivered to the contents of a platinum crucible by a bunsen or tirrill burner is about 850°C; that by a grid-top burner is about 1000°C; that by a gas-air blast lamp about 1150°C.

Cooling. A dried or ignited precipitate is cooled in a desiccator, which is a jarlike receptacle containing a desiccant, and which, in analytical chemistry, is used principally to allow a heated crucible and its contents to come to room temperature without taking on moisture from the air. Anhydrous calcium chloride, although not a very efficient desiccant, is commonly used for this purpose because of its low cost. Other desiccants in the order of increasing effectiveness are anhydrous barium perchlorate, sodium hydroxide sticks, silica gel, aluminum oxide, anhydrous calcium sulfate, calcium oxide, anhydrous magnesium perchlorate, barium oxide, and phosphorus pentoxide.

Calculations. At least two weighings are required for each analysis—the original sample, and the dried or ignited residue. From these weights, the percentage or proportion of the desired constituent may be calculated.

$$\%A = \frac{\text{wt of residue} \times \text{factor} \times 100}{\text{wt of sample}}$$

The factor is determined from a knowledge of the chemical relationships between the weight of substance A contained in, or equivalent to, a fixed weight of residue of known composition. *See* ANALYTICAL CHEMISTRY; ELECTRODEPOSITION ANALYSIS; QUANTITATIVE CHEMICAL ANALYSIS; STOICHIOMETRY. [S.C.S.]

Gravitation

The mutual attraction between all masses and particles of matter in the universe. Until the seventeenth century, the only evidence of this phenomenon was the gravitational acceleration at the surface of the Earth. Only vague speculation existed that some force emanating from the Sun kept the planets in their orbits. Such a view was expressed by Johannes Kepler (1571–1630), the author of the laws of planetary motion, but Kepler lacked a detailed understanding of the principles of mechanics, and without this no real progress could be made. The formulation of the three laws of motion by Isaac Newton (1643–1727) was therefore an essential preparation for his formulation of the law of gravitation. *See* CELESTIAL MECHANICS; NEWTON'S LAWS OF MOTION; PLANET.

Newton's law of gravitation. Newton's law of universal gravitation states that every two particles of matter in the universe attract each other with a force which acts in the line joining them and whose intensity varies as the product of their masses and inversely as the square of the distance between them.

Newton made this bold generalization after contemplating gravity at the Earth's surface, Kepler's laws of planetary motion, and the Moon's orbit around Earth. The gravitational force F exerted by two particles with masses m_1 and m_2 separated by a distance r is, by Newton's law,

$$F = \frac{Gm_1m_2}{r^2}$$

in which G is called the constant of gravitation. In order to evaluate the attraction between two finite masses, the components of the force must be integrated over all the mass elements of the two bodies. Bodies with spherically symmetrical distribution of mass attract each other as if all their mass were concentrated at their respective centers. For bodies with arbitrary distribution of mass, the simple formula so obtained is an approximation which is the more nearly exact the greater the distance between the centers of mass compared with the dimensions of the bodies.

Let E be the mass of Earth, assumed to be spherically symmetrical with radius R. Then the force exerted by Earth on a small mass m near Earth's surface is given by

$$F = \frac{GEm}{R^2}$$

and the acceleration of gravity g by

$$g = \frac{GE}{R^2} \qquad (1)$$

Let a be the mean distance of the Moon from Earth, M the Moon's mass, P the Moon's period of revolution. If the motions in the Earth-Moon system are considered to be unaffected by external forces (principally those caused by the Sun's attraction), Kepler's third law applied to the Earth-Moon system would be

$$\frac{4\pi^2}{P^2}a^3 = G(E + M) \qquad (2)$$

Equations (1) and (2), on elimination of G, give

$$g = 4\pi^2 \frac{E}{E + M} \cdot \frac{a^3}{R^2} \cdot \frac{1}{P^2} \qquad (3)$$

Now the Moon's mean distance from Earth is $60.267\ R = 3.8110 \times 10^{10}$ cm; the sidereal period of revolution is 27.322 days $= 2.3606 \times 10^6$

These data give, with $E/M = 81.35$,

$$g = 39.478 \times 0.98785 \times 3632.1 \times \frac{3.844 \times 10^{10}}{5.5724 \times 10^{12}}$$

$$= 977.1 \text{ cm sec}^{-2}$$

This computed value is close to the observed value of the acceleration of gravity at the surface of Earth (*see* TERRESTRIAL GRAVITATION).

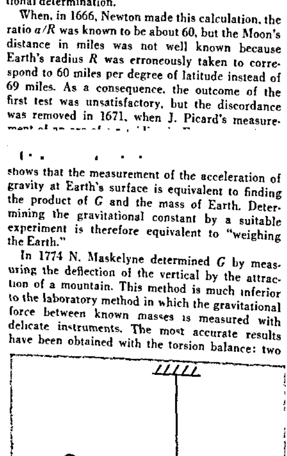
The calculation corresponds, in essence, though not in detail, to the calculation that Newton made in order to verify that the gravitational attraction experienced at the surface of Earth, diminished according to the square of the distance from Earth's center, agrees with the force required to keep the Moon in its orbit.

It is of interest to note that in modern times, Eq. (3), in a modified form with appropriate refinements to allow for Earth's oblateness and the solar attraction of the Earth-Moon system, yields a computed value of the Moon's distance with an accuracy that surpasses that of the direct observational determination.

When, in 1666, Newton made this calculation, the ratio a/R was known to be about 60, but the Moon's distance in miles was not well known because Earth's radius R was erroneously taken to correspond to 60 miles per degree of latitude instead of 69 miles. As a consequence, the outcome of the first test was unsatisfactory, but the discordance was removed in 1671, when J. Picard's measurement of an arc of [illegible]

[illegible] shows that the measurement of the acceleration of gravity at Earth's surface is equivalent to finding the product of G and the mass of Earth. Determining the gravitational constant by a suitable experiment is therefore equivalent to "weighing the Earth."

In 1774 N. Maskelyne determined G by measuring the deflection of the vertical by the attraction of a mountain. This method is much inferior to the laboratory method in which the gravitational force between known masses is measured with delicate instruments. The most accurate results have been obtained with the torsion balance: two



The torsion balance.

small spheres, each of mass m, are connected by a light rod, suspended in the middle by a thin wire. The deflection caused by bringing two large spheres of mass M near the small ones on opposite sides of the rod is measured, and the force is evaluated by observing the period of oscillation of the rod under the influence of the torsion of the wire (see illustration).

This is known as the Cavendish experiment, in honor of H. Cavendish who achieved the first reliable results by this method in 1797–1798. More recent determinations of G with the torsion balance in various forms have been made by C. V. Boys in England, C. Braun in Austria, and by P. R. Heyl (later in collaboration with P. Chrzanowski) in the United States. On the whole, the determinations are in good agreement. They yield:

Constant of gravitation $G = 6.67 \times 10^{-8}$ cgs units
Mass of Earth $= 5.98 \times 10^{27}$ g
Mean density of Earth $= 5.52 \text{ g cm}^{-3}$

The uncertainty of these three numbers is probably about one-half unit of the last place given.

Gravity. This term denotes the acceleration of gravity at the surface of a planet or other celestial body. A rotating planet is flattened at the poles to a form such that the observed acceleration of gravity at any point (that is, the planet's attraction diminished by the centrifugal acceleration of the rotation) is perpendicular to the surface.

A planet's oblateness depends on the ratio ϕ of the centrifugal to the gravitational acceleration at the planet's equator and also on the distribution of mass in the planet's interior. For a homogeneous planet, the oblateness ϵ (equatorial minus polar diameter divided by equatorial diameter) is $\epsilon = 5\phi/4$; for a body with extreme concentration of mass at the center, it is $\epsilon = \phi/2$. Among the planets in the solar system for which reliable values of ϕ are available, the ratio ϵ/ϕ varies from 1.14 for Mars and 0.98 for Earth to 0.77 for Jupiter and 0.69 for Saturn.

As a consequence of the rotational flattening and the varying centrifugal acceleration in different latitudes, the observed acceleration g at sea level increases from the equator to the poles. For an elevation h above sea level, the acceleration of gravity diminishes by the amount $\Delta g = -2gh/R$, provided that the elevation is small compared with R and that the effect of the planet's oblateness can be ignored. *See* CENTRIFUGAL FORCE; FREE FALL.

Gravitational potential energy. If two particles with masses m_1 and m_2 are a distance r apart, and if the distance is increased to $r + \Delta r$, the work done against the gravitational attraction is $Gm_1m_2\Delta r/r^2$. If the distance is increased by a finite amount, say from r_1 to r_2, the work done is

$$W'_{r_1,r_2} = Gm_1m_2 \int_{r_1}^{r_2} \frac{dr}{r^2}$$

$$= Gm_1m_2 \left(\frac{1}{r_1} - \frac{1}{r_2}\right)$$

and, if $r_2 \to \infty$

$$W_{r_1,\infty} = G\frac{m_1 m_2}{r_1}$$

Hence, if the gravitational potential energy for the system of two particles is (arbitrarily) put equal to zero for infinite distance between the particles, then the gravitational potential energy of the system when the distance between the particles is r is

$$-U = -\frac{Gm_1m_2}{r}$$

Similarly, for a system of n particles with masses $m_1, m_2, \ldots, m_n$, the gravitational potential energy $-U$ is the negative of the work required to bring about an infinite separation:

$$-U = -G\sum_{k>j}\frac{m_j m_k}{r_{j,k}}$$

If T represents the kinetic energy of a system of particles, conservation of energy is expressed by the energy integral, also called the vis viva integral

$$T - U = \text{constant}$$

See ENERGY.

Possible nature of gravitation. In the foregoing parts of this article, gravitational attraction has been treated as the instantaneous action of one body upon another at a distance through a vacuum. Newton himself declared this assumption "an absurdity."

The merit of the ad hoc adoption of the Newtonian law of universal gravitation within the framework of Newtonian mechanics is that it permits a simple precise formulation of gravitational problems that works remarkably well. After more than two centuries of increasingly refined application to the solution of problems concerning the motions of bodies in the solar system, only a few deviations from classical gravitational theory had become clearly established. In fact, at the beginning of the twentieth century the Moon's motion and the excess motion of the perihelion of Mercury presented the only well-established deviations from gravitational theory as applied to the motions of bodies in the solar system.

The supposed fluctuations in the Moon's motion were subsequently found to be adequately accounted for by *irregularities in the rate* of rotation of Earth, and are therefore no contradiction of gravitational theory.

The excess motion of the perihelion of Mercury was first established by U. J. J. Leverrier and subsequently confirmed by Simon Newcomb. It amounts to approximately 43 sec of arc per century. Attempts to account for this excess motion on the basis of classical gravitational theory by various postulates were not successful. Newcomb, in 1895, adopted as a provisional solution the hypothesis ascribed to Asaph Hall that the exponent 2 in the inverse square law should actually be $2 + \delta$, and found that $\delta = 1.574 \times 10^{-7}$ would yield the observed motion of the perihelion of Mercury without serious contradictions in other observed quantities.

The discordance was removed by Albert Einstein's *general theory of relativity in 1916.* Einstein's concepts replaced the action at a distance of the classical theory by treating the phenomenon of gravitation as a consequence of the geometric properties of space time, the geometry of space being affected by the presence of gravitating matter. For the foundation of Einstein's theory of gravitation and its observational tests, *see* RELATIVITY; SPACE-TIME.

Soon after the appearance of the general theory of relativity, the need was felt for a more comprehensive theory that would establish greater unity between gravitational and electromagnetic phenomena. For a summary of attempts in this direction, *see* UNIFIED FIELD THEORIES.

The discoveries of the numerous fundamental particles in the twentieth century have raised a new question concerning the nature of gravitation: does gravitation alone remain as a geometric property of space, while electromagnetic and nuclear forces are ascribed to local interactions between particles? The alternative appears to be that gravitational interaction also should be reducible to local interactions The graviton is a theoretically deduced particle postulated as the quantum of the gravitational field. *See* GRAVITON; *see also* QUANTUM FIELD THEORY.

Against this similarity is the fact that gravitational interaction is much weaker than these other forces by a factor of the order 5×10^{-40} if gravitational interaction between proton and electron is compared with their electrostatic interaction

Gravitational and inertial mass. In the equations of motion of Newtonian mechanics, the mass of a body appears as inertial mass as a factor of the acceleration, and as gravitational mass in the expression of the gravitational force (*see* MASS). In the Newtonian theory, the equality of inertial mass and gravitational mass is a coincidence observationally confirmed by the fact that the acceleration of gravity at the surface of Earth is the same for all bodies regardless of their composition. In Einstein's law of gravitation, the equality is no longer a coincidence, since inertia and gravitation are unified: inertia of a body is its interaction with the general character of the space-time in which it lies. In the philosophy of E. Mach, inertia of one test body in an otherwise empty world is absent; it requires the presence of other matter. According to Mach's principle, the inertial force is the gravitational interaction with distant matter.

Velocity of propagation. On the basis of relativistic theory, it is difficult to conceive that gravitational attraction should be propagated with a velocity greater than that of light. Early attempts to modify the Newtonian equations by requiring a finite speed of propagation led to negative results. Observational contradictions arose unless the speed of propagation were of the order of 10^7 times of light. A satisfactory solution was obtained

after the discovery of the gravitational waves of the general theory of relativity. This theory removes the objections to the older treatments; the velocity of propagation of gravitational attraction equals the velocity of light, and the observational contradictions are no longer present.

[D.BR.]

Bibliography: R. H. Dicke, Gravitation, an enigma, *Am. Scientist*, 47(1):25–40, 1959; G. P. K[illegible] (ed.), *The Solar System*, vol 2, 1954; E. T. [illegible] *the Aether* [illegible] taker, *From Euclid to Eddington, a Study of Conceptions of the External World*, 1949.

Graviton

A theoretically deduced particle postulated as the quantum of the gravitational field. According to Einstein's theory of general relativity, accelerated masses (or other distributions of energy) should emit gravitational waves, just as accelerated charges emit electromagnetic waves. And according to quantum field theory, such a radiation field should be quantized, that is, its energy should appear in discrete quanta, called gravitons, just as the energy of light appears in discrete quanta, namely photons.

The properties of the graviton follow from the properties of the classical gravitational field, namely: its rest mass and charge are zero (like the photon); it has spin 2 in units of $h/2\pi$, where h is Planck's constant, and is therefore a boson, that is, a particle which obeys Bose-Einstein statistics. Because of its vanishing rest mass, its spin is restricted to be parallel to its motion, so that a graviton has only two independent spin states (again like a photon).

The observation of gravitational radiation, not to mention the verification of its quantized nature, has not yet (1959) been accomplished. It will be exceedingly difficult, because matter is very weakly coupled to the gravitational field (the gravitational force between an electron and a proton is only 10^{-39} times the electrical force between them), so that the rate of emission and absorption of gravitational radiation is very small. *See* ELEMENTARY PARTICLE; GRAVITATION; QUANTUM FIELD THEORY; RELATIVITY.

[C.J.G.]

Bibliography: G. Wentzel, *Quantum Theory of Fields*, 1949.

Gravity

The gravitational attraction at the surface of a planet or other celestial body. The quantity g is often referred to simply as "gravity" or "the force of gravity" of Earth, both of which are incorrect. The force of gravity means the force with which a celestial body attracts an object, that is, the weight of the object. The letter g represents the acceleration caused by the gravitational force and, of course, has the dimensions of acceleration. *See* GRAVITATION; TERRESTRIAL GRAVITATION; *see also* ANTIGRAVITY; SPACE BIOLOGY.

[D.BR.]

Gray

The term used of all the intermediate colors in the series of achromatic colors from the extreme of white to the other extreme of black (*see* BLACK, WHITE). Gray may result from a mixture of two complementary colors, from a mixture of all primary colors, or from a fairly uniform mixture of lights of all wavelengths throughout the visible spectrum (see COLOR; COLOR VISION). Grayness is relative; a light gray is an achromatic color that is lighter than its surroundings, while a dark gray is simply darker than its surroundings. Thus, the same surface may be called light or dark gray when carried from one situation to another (*see* VISION). A gray color is one of minimum saturation; it corresponds to zero "excitation purity."

[L.A.R.]

Gray body

An energy radiator which has a black-body energy distribution, reduced by a constant factor, throughout the radiation spectrum or within a certain wavelength interval (*see* BLACK BODY). The designation "gray" has no relation to the visual appearance of a body but only to its similarity in energy distribution to a black body. Most metals, for example, have a constant emissivity within the visible region of the spectrum and thus are gray bodies in that region. The gray body concept allows the calculation of the total radiation intensity of certain substances by multiplying the total radiated energy (as given by the Stefan-Boltzmann law) by the emissivity. The concept is also quite useful in determining the true temperatures of bodies by measuring the color temperature.

For a discussion of the Stefan-Boltzmann law and color temperature, *see* HEAT RADIATION.

[H.G.S.; F.J.W.]

Graywacke

An argillaceous sandstone ("dirty" sandstone) that is characterized by an abundance of unstable mineral and rock fragments and a fine-grained clay matrix binding the larger sand-size detrital fragments. Perhaps the most important characteristic is the clay matrix, which constitutes a minimum of 15% of the rock and may, in extreme cases, make up almost half of the bulk.

Mineral composition. The unstable mineral fragments include feldspar, augite, hornblende, serpentine, biotite, chlorite, and magnetite. Rock fragments include many varieties of low- and high-grade metamorphic rocks (phyllite, slate, quartzite, and granulite) as well as aphanitic (microscopically crystalline) volcanic rocks, typically greenstones and spilitic rocks. The clay matrix is composed of a mixture of micaceous minerals, with biotite and chlorite predominating over muscovite (and illite). Kaolinite, abundant in arkoses, is almost completely absent in the graywackes. Some graywackes have the appearance of water-laid tuffs and may grade into them. These are

quartz poor and have abundant hornblende and volcanic rock fragments.

Texture. The sorting of graywackes is poor. Inevitably the presence of a clay matrix contributes to the poor sorting. Thus the size distribution curve is skewed towards the fine sizes. Even when the matrix is not considered, the larger detrital grains are poorly sorted. The grains tend to be sharp and angular and may show little or no evidence of abrasion. Many of the grains of metamorphic derivation have low sphericity, either being roughly rod-shaped (quartz, amphiboles, and pyroxenes) or flat (micas). Precipitated mineral cements such as silica or calcite are absent.

Structure. The sedimentary structures of the graywackes differ greatly from those of orthoquartzites and arkoses. Bedding is thin and individual laminae may persist for hundreds of yards. The individual laminae are hard and homogeneous and tend to break with subconchoidal fracture. Graded bedding—that is, repetitive sequences of beds, each bed containing a range of particle sizes grading upward from coarse to fine—is a significant feature. Normally the graded bedding is of the type in which the upward gradation from coarse to fine is the result of gradual diminution of the coarsest size upward, while fine sizes remain more or less constant throughout the entire interval. Crossbedding is absent in most beds; if present, it is only a few inches thick. A group of bedding-plane structures, variously called flow casts, groove casts, and lobate rill marks are commonly found in this rock type. Penecontemporaneous deformation features, small folds and thrusts formed shortly after deposition, may often be found in graywackes. Most common is convolute bedding, a plastic deformation of a bed between undeformed beds.

Graywackes are found in association with slates, pillow lavas, cherts, and greenstones. Their occurrence is restricted to geosynclinal areas Almost all ages are known, from Precambrian to Tertiary *See* GEOSYNCLINE.

Origin. The genetic significance of graywackes is revealed by the clay matrix, the presence of a large amount of unstable mineral and rock fragments, and the distinctive sedimentary structures found in them. The presence of a large amount of clay matrix implies special sedimentation conditions under which the depositing current does not sort fine from coarse material. The investigation of turbidity (density) currents as agents of transportation has strengthened the hypothesis that the graywackes probably represent a large group of rocks that have been formed by such currents. Further corroboration of this hypothesis is given by the prevalence of graded bedding and other sedimentary structures of the graywackes. Deep-water origin for graywackes has been argued by some geologists, primarily because in deep water normal currents would not be present to sort the material whereas turbidity currents could be the dominant agent of transport. The unstable mineral and rock fragments imply, as in the arkoses, tectonically active source areas. The penecontemporaneous deformation and association with pillow lavas implies tectonic activity in the basin of deposition as well. *See* ARKOSE; ORTHOQUARTZITE; SANDSTONE; SEDIMENTARY ROCKS; SUBGRAYWACKE; TURBIDITY CURRENT. [R.S.]

Great circle, terrestrial

A circle or near-circle described on the earth's surface by a plane passing through the center of the earth. The most important such circle is the Equator, cut by the plane which bisects, and is perpendicular to, the earth's axis. Planes through the poles cut the earth along meridians. The described forms are not quite circular, being slightly flattened ellipsoids, because the equatorial diameter is 1/298.3 times larger than the polar diameter (*see* GEOGRAPHY, MATHEMATICAL). All other great circles have shapes between those of the equator and the meridians. Even the equatorial circle and the ellipsoids have minor irregularities; to find the exact shape of the geoid was a major task of the International Geophysical Year. *See* GEODESY.

Segments of great circles extending from beginning to terminus of proposed global travel make the shortest routes on the earth's surface. Ships sailing for short distances, however, follow compass directions, also called rhumb lines, or loxodromes. Except where they coincide with meridians or the equator, loxodromes are not segments of great circles. In transoceanic voyages, distance can be shortened by following great-circle routes called orthodromes. These are laid out on gnomonic charts which show every orthodrome as a straight line. Ships, however, navigate by compass; thus or-

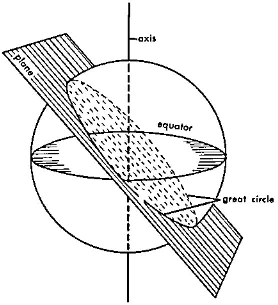


Diagram of a great circle described by a plane through the center of the earth.

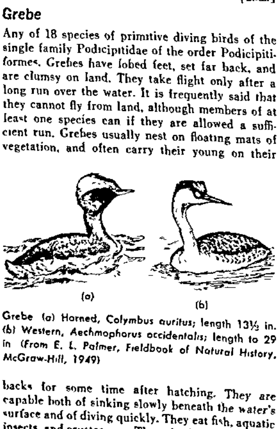
thodromes are transferred to a Mercator chart on which compass directions appear as straight lines; steering is actually by a series of short segments of compass directions. The longer the distance, the greater the deviation between loxodromes and orthodromes. For instance, the compass direction east leads from New York to near Naples, but the shorter great-circle route can be demonstrated by stretching a string taut between the two places on a terrestrial globe. Marine routes would have to deviate to avoid land, but aeronautical routes might more closely approximate true great-circle travel. Intercontinental and transcontinental air routes are laid out as nearly as possible to true great circles. As an example, the direct air route from Panama to Tokyo passes near Houston, Seattle, Anchorage, and the Aleutian Islands. *See* CARTOGRAPHY; MAP PROJECTIONS; NAVIGATION.

Great circles are used in various technical fields. Radio waves propagate along great circles from transmitting to receiving stations. This is utilized in loran and similar navigational systems. Great-circle directions are also important in seismology, ballistics, intercontinental missiles, satellites, and like technical aspects of travel and transmission.

[E.RZ.]

Grebe

Any of 18 species of primitive diving birds of the single family Podicipitidae of the order Podicipitiformes. Grebes have lobed feet, set far back, and are clumsy on land. They take flight only after a long run over the water. It is frequently said that they cannot fly from land, although members of at least one species can if they are allowed a sufficient run. Grebes usually nest on floating mats of vegetation, and often carry their young on their



(a) (b)

Grebe (a) Horned, *Colymbus auritus*; length 13½ in. (b) Western, *Aechmophorus occidentalis*; length to 29 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

backs for some time after hatching. They are capable both of sinking slowly beneath the water's surface and of diving quickly. They eat fish, aquatic insects, and crustaceans. They also have the strange habit of swallowing feathers, sometimes in quantity.

The species common in the eastern United States, the pied-billed grebe, *Podilymbus podiceps*, is frequently called the hell diver. *See* PODICIPITIFORMES.

[J.D B.]

Green snake

Either of two small snakes of the genus *Opheodrys*. The rough green snake, *O. aestivalis*, has keeled scales and is of a uniform light green above and whitish or yellowish below. It is found over most of the United States east of the Great Plains, and is a fairly common arboreal snake; its coloration makes it difficult to detect in such a habitat. It is entirely insectivorous. Most adults are less than 18 in. long, although it has been known to grow to 39 in.

The smooth green snake, *O. vernalis*, is similar, but its scales are smooth rather than keeled. Its back is light green and the belly is paler green to yellowish. It is found in the northeastern United States, across North Dakota and Kansas, and southward in the mountains to southern New Mexico. It usually frequents grassy, hilly places, and commonly takes refuge under stones. It is usually less than 15 in. long, but is known to grow to 26 in. in length at times. *See* SQUAMATA.

[J.D B.]

Green's theorem

A term used variously in mathematical literature to denote either (1) the Gauss divergence theorem $\iiint \nabla \cdot \mathfrak{F}\, dV = \iint \mathfrak{F} \cdot \mathbf{v}\, dS$ or some one of several forms or immediate consequences of this theorem or (2) the plane case of Stokes' theorem. *See* GAUSS' THEOREM; STOKES' THEOREM.

The variation of the Gauss divergence theorem in which $\mathfrak{F}$ is the product $u\,\nabla v$ of a scalar function and the gradient of a scalar function is sometimes called Green's *first identity*. In this case,

$$\nabla \cdot \mathfrak{F} = \nabla \cdot (u\,\nabla) = u\,\nabla^2 v + \nabla u \cdot \nabla v$$

and Gauss' theorem assumes the form

$$\iiint u\,\nabla^2 v\, dV + \iiint \nabla u \cdot \nabla v\, dV = \iint u\, \partial v/\partial n\, dS \tag{1}$$

Here $\partial v/\partial n = \nabla v \cdot \mathbf{v}$ and is therefore the directional derivative of v in the direction of $\mathbf{v}$, the unit normal to the surface; thus n may be taken to be arc length along any curve of class C' which is tangent to $\mathbf{v}$ at the point in question. The derivative $\partial v/\partial n$ is a function of the surface coordinates.

By interchanging u and v in Eq. (1) with subsequent subtraction of the new form from the original, there results the relationship

$$\iiint u\,\nabla^2 v\, dV - \iiint v\,\nabla^2 u\, dV = \iint (u\, \partial v/\partial n - v\, \partial u/\partial n)\, dS \tag{2}$$

… as well as

…ssociated

… $= v$ and

$u = 1$ in Eq. (1), thus

$$\iiint u\,\nabla^2 u\, dV + \iiint \nabla u \cdot \nabla u\, dV = \iint u\, \partial u/\partial n\, dS \tag{3}$$

$$\iiint \nabla^2 v\, dV = \iint \partial v/\partial n\, dS \tag{4}$$

The formulas (1)–(4) provide powerful devices for the investigation of differential equations involving the Laplacian operator ∇^2, such as Poisson's equation and the wave equations. *See* CALCULUS OF VECTORS.

[H V C.]

Greenhouse effect, terrestrial

The earth's atmosphere acts as the glass walls and roof of a greenhouse in trapping heat from the sun. Like the greenhouse, it is largely transparent to solar radiation, but it strongly absorbs the longer-wavelength radiation from the ground. Much of this long-wave radiation is reemitted downward to the ground, with the paradoxical result that the earth's surface receives more radiation than it would if the atmosphere were not between it and the sun.

The absorption of long-wave (infrared) radiation is effected by small amounts of water vapor, carbon dioxide and ozone in the air and by clouds (*see* ATMOSPHERE; OZONE; RADIATION, TERRESTRIAL). Clouds actually absorb about one-fifth of the solar radiation striking them, but, unless they are extremely thin, they are almost completely opaque to infrared radiation. The appearance even of cirrus clouds after a period of clear sky at night is enough to cause the surface air temperature to increase rapidly by several degrees because of radiation from the cloud.

The greenhouse effect is most marked at night, and usually keeps the diurnal temperature range below 20°F. Over dry regions such as New Mexico and Arizona, however, where the water-vapor content of the air is low, the atmosphere is more transparent to infrared radiation, and cool nights may follow very hot days. [L.D.K.]

Greenockite

A mineral having composition CdS and crystallizing in the hexagonal system (dihexagonal pyramidal class). Pyramidal crystals are rare and greenockite usually occurs as earthy coatings with resinous luster and yellow-to-orange color. There is good prismatic cleavage; the hardness is 3 (Mohs scale) and specific gravity is 4.9. Greenockite and wurtzite, ZnS, are isostructural and a complete solid-solution series exists between the two minerals. Although greenockite is the most common cadmium mineral, no deposits of it are sufficiently large to warrant mining it solely as a source of cadmium. It is commonly associated with sphalerite and thus the supply of cadmium comes as a by-product from the treating of zinc ores. *See* CADMIUM; SPHALERITE. [C.S.HU.]

Gregarinidia

A subclass of the class Telosporidia. These protozoans occur principally as extracellular parasites in the digestive tracts and body cavities of invertebrates. Their spores are formed directly by the zygote. There are two orders, the Schizogregarinida whose life cycle embraces both schizogony and sporogony, and the Eugregarinida which increase only by sporogony. The most familiar gregarines belong to the latter order and are represented by two types: (1) cephaline, whose trophozites (sporadins) are divided into an anterior protomerite and a larger posterior deutomerite by a transverse septum, and (2) acephaline, which lack the septum. *See* EUGREGARINIDA; SCHIZOGREGARINIDA; TELOSPORIDEA. [E.R.BE.]

Greisen

A pneumatolytically altered granite consisting of mainly quartz and a light green (white) mica; the feldspars and biotites of the original granite have been replaced more or less completely by muscovite, lithium mica, kaolinite, and tourmaline. Other substances usually introduced during alteration are topaz, apatite, fluorite, and ores of iron, and in some places ores of tungsten and tin. *See* PNEUMATOLYSIS.

Greisen occurs in bands and veins intersecting granite. The bands are rarely more than 2 ft thick with indefinite margins that grade into unaltered granite They were evidently formed by flow of vapors and gases through fissures.

In principle the formation of greisen is easily understood, being a typical example of pneumatolytic metasomatism. A crystallizing granitic magma gives off gases containing silicon, Si; fluorine, F; boron, B; lithium, Li; and often tin, Sn; or tungsten, W. They cause a complete alteration of the adjacent solid rock, be it the granite itself or its country rock. The usual, completely altered product is a rock consisting only of quartz and mica (that is, greisen). In limestone and dolomite, fluorite often accompanied by humite will develop instead of topaz, and instead of tourmaline, the boron-bearing mineral axinite may develop. Datolite also may occur. *See* METASOMATISM.

Tin ore deposits (cassiterite, SnO_2) are usually accompanied by greisen and [illegible]

$$\underset{\text{Tin(IV) fluoride}}{SnF_4} + \underset{\text{Water}}{2H_2O} \rightarrow \underset{\text{Cassiterite}}{SnO_2} + \underset{\text{Hydrofluoric acid}}{4HF}$$

where the liberation of hydrofluoric acid accounts for the strong alteration of the adjacent rocks. [T.F.W.B.]

Grignard reaction

A reaction between an alkyl or aryl halide and magnesium metal in a suitable solvent, usually absolute ether (R = alkyl or aryl; X = chlorine, bromine, or iodine):

$$RX + Mg \xrightarrow{\text{Ether}} RMgX$$

The organomagnesium halide produce[illegible]

[illegible] prepared [illegible] of alkyl and aryl halides. In general, alkyl chlorides, bromides, and iodides, an aryl bromides and iodides react readily. A few ides, such as aryl chlorides, react very slu

and require specially activated magnesium, modified reaction techniques, the use of high-boiling solvents, and long reaction periods. It has recently been found that vinylmagnesium halides can be prepared using tetrahydrofuran as the solvent. Hydrocarbons containing two or more halogen atoms will occasionally undergo normal reactions. For example, tetramethylene dibromide reacts with magnesium to give tetramethylenedimagnesium bromide in good yield:

$$BrCH_2CH_2CH_2CH_2Br + 2Mg \rightarrow BrMgCH_2CH_2CH_2CH_2MgBr$$

Frequently, however, the presence of more than one halogen atom produces halogen elimination or coupling.

The exact structure of the Grignard reagent is not yet known with certainty. Evidence indicates, however, that most Grignard reagents are dietherates of organomagnesium halides, and have the empirical formula $RMgX{\cdot}2(C_2H_5)_2O$. Thus, ether is not only a solvent for the reaction but an integral part of it. It is thought that Grignard reagents in solution actually consist of an equilibrium mixture:

$$2RMgX \rightleftharpoons R_2Mg + MgX_2$$

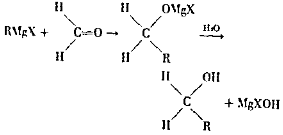
Furthermore, all components of the equilibrium are capable of ionization, although the actual degree of dissociation must be very small:

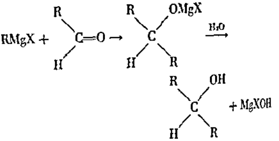
$$RMgX \rightleftharpoons R^- + MgX^+$$
$$R_2Mg \rightleftharpoons R^- + RMg^+$$
$$MgX_2 \rightleftharpoons X^- + MgX^+$$

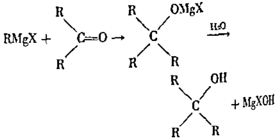
Probably the most common synthetic use of the Grignard reaction involves the reaction of Grignard reagents with carbonyl compounds, followed by hydrolysis to produce a variety of products containing new carbon-carbon bonds. For example, the reaction of a Grignard reagent with carbon dioxide at low temperatures produces salts of carboxylic acids, from which the acids can be readily isolated:

$$RMgX + CO_2 \rightarrow RCOOMgX \xrightarrow{H_2O} RCOOH + MgXOH$$

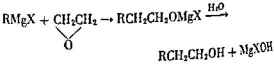
Addition of Grignard reagents to the carbonyl group of many aldehydes and ketones produces primary, secondary, or tertiary alcohols:





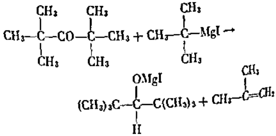


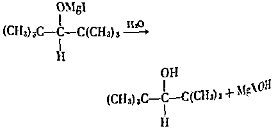
The reaction of Grignard reagents with carbon dioxide, aldehydes, and ketones serves as a convenient means of lengthening the carbon chain. Further, a hydrocarbon chain can be increased by two carbon atoms by the reaction of a Grignard reagent with ethylene oxide:



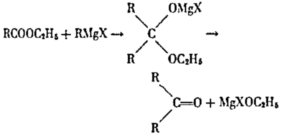
The resulting alcohol can be converted to the corresponding halide, and the process repeated.

In certain instances in which the carbonyl group is sterically hindered, the Grignard reagent serves as a reducing agent. An example is the reaction of di-*tert*-butylketone with *tert*-butylmagnesium iodide to yield di-*tert*-butylcarbinol and isobutylene:





Grignard reagents react with esters to produce ketones initially:



The reaction as a rule does not stop at this stage, however, since additional Grignard reagent reacts with the ketone to yield a tertiary alcohol as illustrated above.

Acid chlorides and anhydrides react in an analogous manner to produce ketones which react further to give tertiary alcohols. A valuable modification of this procedure involves prior reaction of the Grignard reagent with cadmium or zinc chloride. The organocadmium or organozinc compounds are less reactive than those of magnesium and react so slowly with ketones that the conversion of the ketone to the tertiary alcohol can usually be avoided.

When aliphatic Grignard reagents are exposed to air or oxygen, they are oxidized to alkoxides, which on *hydrolysis produce alcohols*. Aromatic Grignard reagents react poorly in this way, and only unsatisfactory yields of phenols can be obtained.

The reaction of Grignard reagents with alkyl halides which contain reactive halogen atoms produces hydrocarbons by a process of alkylation:

$$RMgX + XR \rightarrow R{-}R + MgX_2$$

In a similar manner, the addition of a variety of metal halides, such as silver bromide or cobaltous chloride, tends to couple Grignard reagents:

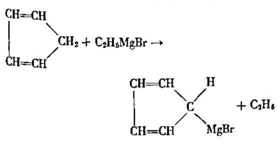
$$2RMgX + 2AgBr \rightarrow R{-}R + 2Ag + 2MgXBr$$

Compounds containing active hydrogen atoms react readily with Grignard reagents. Reaction with water brings about immediate decomposition and formation of the corresponding hydrocarbon:

$$RMgX + H_2O \rightarrow RH + MgXOH$$

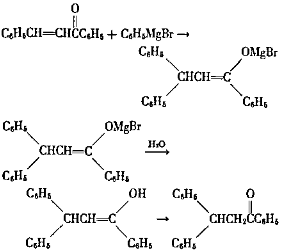
In certain instances, hydrogen attached to carbon is sufficiently acidic to react with Grignard reagents. The reactions of ethylmagnesium bromide with acetylene and with cyclopentadiene are examples:

$$2C_2H_5MgBr + HC{\equiv}CH \rightarrow 2C_2H_6 + BrMgC{\equiv}CMgBr$$



These reactions enable formation of additional Grignard reagents which could otherwise not be prepared by conventional methods.

Although Grignard reagents do not add to carbon-carbon double bonds except when these bonds are very highly activated, they frequently add to conjugated systems. For example, the addition of *phenylmagnesium bromide to benzalacetophenone* occurs almost exclusively in a conjugate manner:



Grignard reagents react with a variety of nitrogen- and sulfur-containing compounds, such as imines, nitriles, and sulfoxides. Furthermore, the reaction of Grignard reagents with halides of boron, phosphorus, silicon, and tin is a very convenient procedure for producing organometallic compounds of these elements. An example is the preparation of tetraphenylsilane from phenylmagnesium bromide and silicon tetrachloride:

$$4C_6H_5MgBr + SiCl_4 \rightarrow (C_6H_5)_4Si + 4MgBrCl$$

See ORGANOMETALLIC COMPOUND. [M.D.R.]

Bibliography: M. S. Kharasch and O. Reinmuth, *Grignard Reactions of Nonmetallic Substances*, 1954.

Grinding

The process of removing metal by the cutting action of a solid, rotating, grinding wheel. The abrasive grains of the wheel perform a multitude of minute machining cuts on the workpiece. Although grinding is sometimes used as the complete machining operation on surfaces, it is generally considered a finishing process used to obtain a fine surface finish and extremely accurate dimensional tolerances. From 0.0005 to 0.020 in. of stock is commonly removed, depending on the type of operation.

Grinding materials. The grinding process is used in machining a wide range of metals in addition to such materials as cemented carbide, marble, stone, and certain ceramics. Because commercial grinding abrasives are many times harder than the metals to be machined, the grinding process

be used on metals too hard to machine otherwise.

Grinding wheels are composed of abrasive grains or grit plus a bonding material. The abrasives most commonly used are silicon carbide (SiC) and aluminum oxide (Al_2O_3) The silcon carbide crystals are very hard and tend to fracture easily. They are used to grind low-tensile strength materials such as aluminum, brass, and copper. They are also used to cut hard, brittle materials such as hard alloys, cemented carbide, gray iron, marble, and stone. Aluminum oxide crystals are tough and tend to resist fracture. In general they are used on high-tensile strength materials such as alloyed steels or wrought iron. *See* ABRASIVE.

Wheel types. Wheels are classified according to type of abrasive (silicon or aluminum), grain size, grade or hardness, structure (referring to density or porosity), and type of bonding material.

Coarse-grained wheels are used for rapid removal of stock; wheels with fine grains cut more slowly but give smoother finishes. Usually the harder the material to be ground, the softer the bond in the wheel should be, while the softer the material to be ground, the harder the bond should be. On soft materials the grit cuts in deeply and therefore requires a strong bonding material to hold it. On hard substances the grit dulls rapidly and a soft bond allows it to tear out and expose new cutting edges. The correct wheel for the job should hold the grain in place until it becomes dull.

Excessive speed may actually hinder cutting; operation of a wheel above its recommended limits can be dangerous.

Wheel care Grinding wheels are kept in proper condition by truing and dressing Truing a wheel means actually to cut it, usually by means of a diamond cutting tool, for the purpose of making it concentric about its axis. Dressing, which opens or reconditions the grinding surface of a wheel to afford maximum cutting qualities, may not necessarily give concentricity. Dressing may be done with steel cutters, abrasive dressing wheels, or by crush dressing. The latter is commonly used to produce a desired shape to the face of the wheel. Crush grinding forces a rotating hard iron wheel of the desired shape into the grinding wheel; the iron wheel crushes away the grain of the grinding wheel until the two wheels mate.

Grinding fluids or coolants are applied to the grinding point to dissipate the heat generated and also to trap the abrasive dust. Although grinding oils are used, water emulsions containing soda or alkali are more common. Plain water could be used except for the problem of rust. Frequently the coolant is filtered to remove abrasive compound and chips.

Grinding machines. Grinding machines are designed to hold a workpiece rigidly and to feed it smoothly through the cutting path of a rotating grinding wheel. These machines must be constructed with sufficient accuracy to produce the finely finished surfaces and close dimensional tolerances expected. Where tolerances are less critical, as in removing burrs or in sharpening a cutting tool, an operator may hand-hold the workpiece against the side or face of a wheel (Fig. 1).

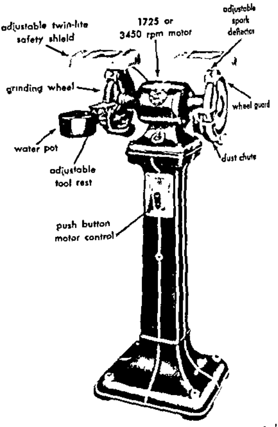


Fig. 1. Power-driven wheels for hand grinding (*Rockwell Mfg. Co.*)

Grinding machines may be listed under four main classifications. They are cylindrical grinders, surface grinders, internal grinders, and special grinders. Grinding machines used for production purposes may be listed under the same four classifications.

Grinding machine operations come under three general classifications. These are cylindrical, surface, and internal grinding. Most special grinding operations are actually variations of these.

Cylindrical grinders. Cylindrical grinding is performed on the peripheries or shoulders of workpieces composed of concentric cylindrical or conical shapes. Examples of such workpieces are shafts, cylinders, rolls, and axles.

Cylindrical grinders rotate the workpiece from a power headstock while it is held between centers, gripped in a jawed chuck, or fastened to a faceplate. Usually, the power headstock and tailstock are held on a worktable which in turn is mounted on the main table. The latter may move longitudinally on ways and thus carry the workpiece past the face of the grinding wheel.

Fig 2. Centerless thread grinding operation. (Landis Machine Co.)

When the work and wheel are moved longitudinally past each other to grind a length greater than the width of the wheel, it is called traverse grinding. If the wheel is advanced directly into the work to form shoulders and contours on the piece, or if the length to be ground is less than the width of the wheel, it is referred to as infeed or plunge grinding.

Centerless grinders. Centerless cylindrical grinders carry the work on a support or blade between two abrasive wheels. One is a regular grinding wheel rotating at normal grinding speed while the second is usually a rubber-bonded abrasive regulating wheel turning in the same direction as the first wheel but at a much slower speed. The regulating wheel does not grind the work but causes it to rotate against the grinding wheel at a uniform speed. Its distance from the grinding wheel determines the finished size of the workpiece.

Because the workpiece is neither rigidly held nor gripped, centerless grinding machines are ideally suited to mass production (Fig. 2). Centerless grinders are made in sizes to handle both large and small workpieces.

With straight, cylindrical workpieces having no interfering shoulders, the regulating wheel may be set to run at a slight angle to the grinding wheel. This causes the rotating workpiece to move longitudinally past the face of the grinding wheel and is termed through feeding.

Workpieces having shoulders and more than one diameter may be ground by infeeding. An end stop keeps the rotating workpiece from moving longitudinally and the regulating wheel is advanced, forcing the work directly against the grinding wheel until the desired diameter is obtained.

End feeding is used when a taper is desired. The grinding wheel, regulating wheel, and support are set at the desired angle and the workpiece is fed in from the side until it reaches the point at which the desired taper and size has been ground on it.

Centerless grinders are designed for both automatic and manual feeds.

Surface grinders. Surface grinding is accomplished by holding one or several workpieces on either a rotary or reciprocating horizontal table and passing them through the cutting path of a rotating grinding wheel. The wheel may be mounted in either a vertical or horizontal plane.

Surface grinders are classed according to the axis of the grinding wheel as horizontal or vertical. Horizontally mounted wheels grind on their peripheries; vertically mounted wheels grind on their circular faces.

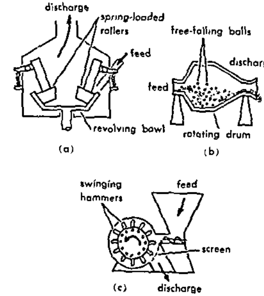
On machines with reciprocating tables, the workpiece is moved back and forth underneath the grinding wheel until the desired dimension and finish are obtained. Where a rotary-table surface grinder is used, enough parts may be mounted on the table to cover its entire working area. As the table rotates, each successive piece passes through the path of the grinding wheel.

Internal grinders. Internal grinding is the process of grinding the surfaces of holes to obtain a close dimensional tolerance or a desired surface finish. Cylindrically shaped holes as well as those with tapered sides may be ground.

Internal grinders vary in the manner in which they hold the workpiece and also in the method of grinding. Some machines may hold the part in a chuck or on a faceplate and rotate it in a fixed position. The rotating grinding wheel reciprocates through the length of the hole and is fed sideways to grind the piece to size. Other types may rotate and also reciprocate the workpiece while the wheel revolves in a fixed position. Large parts, ground on a planetary type grinder, remain stationary with the grinding wheel revolving on its own axis and also rotating around the axis of the hole. Where high production is desired, a centerless-type internal grinder may be employed. *See* HONING; LAPPING; MACHINING OPERATIONS; POLISHING. [A.T.]

Grinding mill

A machine that reduces the size of particles of raw material fed into it. The size reduction may be to facilitate removal of valuable constituents from an ore or to prepare the material for industrial use, as in preparing clay for pottery making or coal for furnace firing. Coarse material is first crushed. The moderate-sized crushings may be reduced further by grinding or pulverizing.



Basic grinding mills. (*a*) Ring-roller mill. (*b*) Tumbling mill. (*c*) Hammer mill.

Grinding mills are of three principal types as illustrated. In ring-roller pulverizers, the material is fed past spring-loaded rollers. The rolling surfaces apply a slow large force to the material as the bowl or other container revolves. The fine particles may be swept by an air stream up out of the mill. In tumbling mills the material is fed into a shell or drum that rotates about its horizontal axis. The attrition or abrasion between particles grinds the material. The grinding bodies may be flint pebbles, steel balls, metal rods lying parallel to the axis of the drum, or simply the larger pieces of the material itself (*see* PEBBLE MILL). In hammer mills, driven swinging hammers reduce the material by sudden impacts.

Depending on the required fineness and uniformity of the finished particles, the discharge may or may not be classified by size. Oversized particles may be returned in a closed circuit to the grinding mill for further reduction. Other variations include grinding the material dry or wet, or in batches or continuously. *See* CLASSIFICATION, MECHANICAL; CRUSHING AND PULVERIZING. [R.M.H.]

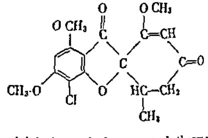
Bibliography: J. H. Perry (ed.), *Chemical Engineers' Handbook*, 3d ed., 1950.

Griseofulvin

An antifungal antibiotic effective in the treatment of ringworm and fungus infections of the skin, hair, and nails. Griseofulvin was originally discovered and isolated by A. E. Oxford and co-workers in England in 1939 as a metabolic product of *Penicillium griseofulvum*. In 1946, P. W. Brian and his co-workers isolated an antibiotic known as curling factor from cultures of *P. janczewskii*, and exposure of *Botrytis allii* to this curling factor resulted in the production of [illegible] that the curling factor antibiotic was identical with griseofulvin described previously by Oxford. It has been subsequently reported to be produced also by *P. urticae*, *P. nigricians*, and *P. raistrickii*.

In 1958, oral griseofulvin was reported as useful in the control of experimental infections of guinea pigs with *Trichophyton mentagrophytes* and *Microsporium canis* and as successful in treatment of cattle with ringworm caused by *T. verrucosum*. In the same year, the first reports of the treatment of infected human beings by the oral administration of griseofulvin appeared. *See* ANTIBIOTIC.

Structure and solubility. Grove and co-workers in 1952 reported the structure of griseofulvin which is given below.



Griseofulvin is a colorless, neutral, thermostable substance which is only very slightly soluble in water. Its em[illegible] and its molecular [illegible] soluble in [illegible] alcohol, 0.1%; chloroform, 4.4%; acetone, 5.0%; and dimethylformamide, 12.0%. In solution, it can withstand autoclaving without loss of potency.

Production. The antibiotic is a metabolic product of various species of *Penicillium* when grown in a standard fermentation medium consisting of an organic nitrogen source, such as corn steep liquor, and a carbon source, such as glucose. The antibiotic, found mainly in the mycelium, is also free in the fermentation broth. It reaches its maximum titer at 9–14 days, and can be extracted with organic solvents.

Antimicrobial activity. The antibiotic is active in vitro at low concentrations against a variety of fungi, particularly against the pathogenic dermatophytes. The minimum inhibitory concentration in broth for a variety of species ranges from 0.1 to 10.0 μg/ml. Griseofulvin is fungistatic rather than fungicidal. It is active against species of *Epidermophyton*, *Microsporum*, and *Trichophyton*, which attack the skin, hair, and nails of man and animals. It is inactive against bacteria, yeasts, and yeast-like molds which cause histoplasmosis and blastomycosis. It has been used in agriculture for the control of plant fungal pathogens, particularly infections of lettuce and tomatoes caused by species of *Botrytis* and *Alternaria*.

Griseofulvin is indicated in the treatment of infections caused by dermatophytic fungi of the hands, feet, fingernails, toenails, glabrous skin, beard, and scalp. Onychomycosis and the following tineas respond to oral therapy: capitis, barbae, cruris, corporis, pedis, tonsurans, and rubrum of

those organisms which cause these conditions, the following are responsive: *Epidermophyton floccosum, Microsporum gypseum, M. canis, M. audouini, Trichophyton metagrophytes, T. rubrum, T. schoenleini, T. sulphureum, T. verrucosum,* and *T. interdigitale.* See DERMATOPHYTOSIS.

Dose and pharmacology. When griseofulvin is administered orally, it is adsorbed from the gastrointestinal tract and deposited in the keratin tissue of newly growing skin, hair, and nails in a concentration sufficient to inhibit the growth of dermatophytic fungi. The fungi remain viable, but are not able to invade the newly growing areas. As the skin, hair, and nails continue to grow and are shed, or cut, they are replaced by normal structure, free from fungi.

Griseofulvin is administered orally in divided daily doses of 1 g per day for adults. As a general rule, therapy should continue until regeneration of new tissue is complete, which usually requires two to four weeks of treatment. Clinical relapse will occur if the medication is not continued until the infecting organism is eradicated. At a level of 1 g per day, improvement is usually noted in four to seven days.

There is complete agreement by investigators that griseofulvin has not shown any toxic properties in man when taken by the oral route. H. Blank (1959) reported normal values in patients with respect to body weight, fasting blood sugar, blood urea nitrogen, alkaline phosphatase, thymol turbidity tests, measurements of the electrolytes in the blood, complete blood counts, urinalyses, and sternal bone-marrow examination. Side effects have occurred which are minimal and transient, such as occasional diarrhea, nausea, and headache. Penicillin-sensitive patients have been treated without difficulty. By the parenteral route of administration of massive doses in rats (2 g/kg), toxicity is observed, but mice survive up to 50 g/kg orally. Interference with mitoses has been observed when 10 times the oral therapeutic dose is administered by parenteral route in rats. See MITOSIS.

[M.J.W.]

Bibliography: H. Blank and F. J. Roth, The treatment of dermatomycoses with orally administered griseofulvin, *A.M.A. Arch. Dermatol.*, 79(3): 259–266, 1959; P. W. Brian, Studies on the biological activity of griseofulvin, *Ann. Botany London,* 13(49):59–78, 1949.

Groin (engineering)

A structure devised to establish, stabilize, or widen a protective beach by trapping littoral drift or by retarding erosion. Groins are usually oriented perpendicular to the shore, and their length depends upon the extent of protection desired. A system of short groins is often more efficient than one long one. Groins may be permeable or impermeable, high or low, fixed or adjustable, and may be constructed of timber, steel, stone, concrete, and other materials or combinations thereof. See COASTAL ENGINEERING.

[E.J.Q.]

Grosbeak

Any of several perching birds, mostly American, with short, stout bills. All are related to the sparrows, although they are divided among different families by recent studies.

The pine grosbeak, *Pinicola enucleator,* breeds in the boreal coniferous forests of Europe and North America, and is an irregular winter visitor to the central United States. The male's color is pinkish overlying gray; the female and young are smoky gray.

The rose-breasted grosbeak, *Pheucticus ludovicianus;* length to 8½ in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

The beautiful rose-breasted grosbeak, *Pheucticus ludovicianus,* is a summer resident of the northwestern United States. The male is a brilliant black and white, strikingly marked with a rose-colored breast. The black-headed grosbeak, *P. melanocephalus,* is a related form found from the Great Plains westward.

The male blue grosbeak, *Guiraca caerulea,* is ultramarine blue, with black wings and tail; the female is brown. It breeds all across the southern half of the United States.

The male evening grosbeak, *Hesperiphona vespertina,* is a chunky, short-tailed, yellow and black finch, with prominent white wing patches; the female is primarily silvery gray. This is a bird of the spruce forest, and is a rather erratic winter visitor elsewhere. See FINCH; PASSERIFORMES; SPARROW.

[J.D.B.]

Ground state

In quantum mechanics, the stationary state of lowest energy of a particle or a system of particles. The ground state may be bound or unbound; when bound, its energy generally is a finite amount

than the energy of the next higher or first excited state. In the typical circumstance that the potential energy is zero at infinite separation, the magnitude of the negative ground state energy is the binding energy, that is, the energy required to separate all the particles infinitely. *See* BINDING ENERGY, NUCLEAR; *see also* ENERGY LEVEL (QUANTUM MECHANICS); EXCITED STATE; QUANTUM THEORY, NONRELATIVISTIC; STATIONARY STATE. [E.G.]

Ground water

The water in the zone in which the rocks and soil are saturated, the top of which is the water table. The zone of saturation is the source of water for wells, which provide about one-fifth of the water supplies of the United States. It is also the chief source of the water that issues as springs and seeps and maintains the dry-weather flow of most perennial streams. The saturated zone is the great natural reservoir which absorbs and stores precipitation during wet periods and pays it out slowly during dry periods; it is, therefore, a natural regulating mechanism which tempers the severity of floods and droughts. The amount of ground water stored in the rocks is estimated to be several times as great as that stored in all lakes and reservoirs, including the Great Lakes.

Ground and subterranean water. Unfortunately, many writers have equated ground water with subterranean water. Subterranean water includes all the water beneath the land surface. It may be subdivided into two parts, water in the zone of aeration above the water table and water in the zone of saturation below the water table. Water in the zone of aeration, also called vadose water or kremastic water, is divided into soil water or rhizic water, intermediate vadose water or argic water, and water of the capillary fringe or anastatic water. Water in the zone of saturation, or ground water, is also called *plerotic water.*

Water in the capillary fringe is connected with the zone of saturation and is held above it by capillary forces. The lower part of the fringe may be saturated but is not a part of the zone of saturation because the water is under less than atmospheric pressure and will not flow into a well, although the walls of the well are moist. When the well reaches the zone of saturation, water will begin to enter it and will stand at the level of the water table.

Rock formations capable of yielding significant volumes of water are called *aquifers*. Some wells are artesian; that is, the water rises above the top of the water-bearing bed. This is a special case and will be discussed later. Other wells encounter water above the saturated zone and lose their water if extended through the impermeable bed upon which the water rests. Such bodies of water are said to be perched. This also is a special case.

Subterranean water occurs in a geologic environment, and therefore a knowledge of geology is essential to an understanding of the occurrence of water. For this reason, the study of ground water is sometimes called hydrogeology or geohydrology. The hydrologist is particularly concerned with the number and size of the openings in rocks and soils and the manner in which they are interconnected. Variations in these openings are almost infinitely diverse. Openings are practically absent in some of the igneous rocks. They are numerous but microscopic in clay. They are large and interconnected in many sands and gravels. There are huge caverns and tubes in many limestones and lavas. Their distribution and type are as diverse as the geology itself, so that general statements about them applicable to one area may be incorrect for another.

Openings in rocks. These are of two broad types. Primary are those which existed when the rock was formed, and secondary, those which resulted from the action of physical or chemical forces after the rock was formed. Primary openings are found in sedimentary rocks such as sand and clay and certain kinds of limestone composed of triturated shells. Certain openings in lava are formed at the stage when the lava is partly liquid and partly solid and also are considered primary. Most rocks containing primary openings are relatively young geologically. Those which contain primary openings large enough to carry useful amounts of water are represented, for example, by the seaward-dipping strata of the Atlantic and Gulf Coastal Plains, including the coquina limestone of Florida, the intermontane valleys of the western United States, the glacial deposits of the United States, and the lava rocks of the Pacific Northwest.

Secondary openings are common in older rocks. Sand and gravel that have been cemented by chemical action, limestone indurated by compression or recrystallization, schist, gneiss, slate, granite, rhyolite, basalt and other igneous rocks, and shale generally contain few primary openings; but they all may contain fractures that will carry water. Limestone in particular is subject to solution which, beginning along small cracks, may develop channels ranging from openings a fraction of an inch across to enormous caverns capable of carrying large amounts of water.

Porosity. The property of rocks for containing voids, or interstices, is termed porosity. It is expressed quantitatively as the percentage of the total volume of rock that is occupied by openings. It ranges from as high as 80% in newly deposited silt and clay down to a fraction of 1% in the most compact rocks.

Permeability. This is the characteristic capability of rock or soil to transmit water. The porosity of a rock or soil has no direct relation to the permeability or water-yielding capacity. This capacity is closely related to the size and the degree to which the pores or openings are interconnected. If the pores are small, the rock will transmit water very slowly; if they are large and interconnected, they will transmit water readily. The standard coefficient of permeability used in the hydrologic work of the U.S. Geological Survey is defined as

the rate of flow of water at 60°F, in gallons per day through a cross section 1 ft sq, under a hydraulic gradient of 100% (1 ft of head loss per foot of water travel). Under field conditions, the adjustment to standard temperature is commonly ignored, and permeability is expressed as a field coefficient at the prevailing water temperature.

Transmissibility. Another coefficient that is commonly used, transmissibility, expresses the rate at which water moves through a saturated body of rock. It is expressed as the rate of flow of water at the prevailing temperature, in gallons per day through a vertical strip of aquifer 1 ft wide, extending the full saturated height of the aquifer under a hydraulic gradient of 100%. This coefficient is especially useful for expressing the total yield of an aquifer.

Forces controlling water in rocks. Water moves through permeable rocks under the influence of gravity from places of higher head to places of lower head, that is, from areas of intake or recharge to areas of discharge, such as wells or springs. Water moving through rocks is acted upon also by friction and by molecular forces. The molecular forces are the attraction of rock surfaces for the molecules of water (adhesion) and the attraction of water molecules for one another (cohesion). When wetted, each rock surface is able to retain a thin film of water despite the effect of gravity. In very fine-grained rocks such as clay and fine silt, the interstices may be so small that molecular attraction extends from one side of a pore to the opposite side. Molecular force then becomes dominant, and water moves through the rock only very slowly under the gradients typical of natural conditions.

The amount of water that drains from a saturated rock under the influence of gravity, expressed as a percentage of the total volume of the rock, is called the specific yield. The percentage of water retained in the rock is called the specific retention. Specific yield is often called effective porosity because it represents the pore space that will surrender water to wells and so is effective in supplying water for human use. The term porosity is poorly defined and its use should be discontinued. A part of the water stored in the rocks is held by molecular forces and may have only a small share in supplying springs or wells. This latter portion is of special [illegible] because it

a soil that is relatively impermeable retains much of its water until it is extracted by plants or by evaporation.

Source of ground water. It has been firmly established that the chief means of replenishment of ground water is downward percolation of surface water, either direct infiltration of rainwater or snow-melt or infiltration from bodies of surface water which themselves are supplied by rain or snow-melt. Evidence on the replenishment of ground water is furnished by analysis of data on the downward movement of precipitation through the soil and subsoil, the rise and fall of ground-water levels and spring discharge in response to precipitation and seepage losses from streams, and the slope of the hydraulic gradient from known areas of intake to areas of discharge. Some ground water may originate by chemical and physical processes that take place deep within the earth. Such water is called juvenile water or primary water to indicate that it is reaching the earth's surface for the first time. The available evidence indicates that such water is always highly mineralized. Some water is stored in deep-lying sedimentary rocks and is a relic of the ancient seas in which these rocks were deposited It is called connate water. The total quantity of water from juvenile and connate sources that enters the hydrologic cycle is insignificant when compared with the quantities of water derived from precipitation (meteoric water). It is balanced to some extent by withdrawal of water from the hydrologic cycle by such processes as deposition of minerals that include water in their crystalline structure. *See* HYDROLOGY.

Infiltration. Replenishment of water in the zone of saturation involves three steps: (1) infiltration of water from the surface into the rock or soil that lies directly beneath the surface, (2) downward movement of a part of the water, the part not retained by molecular forces through the zone of aeration, and (3) entrance of this part of the water into the zone of saturation where it becomes ground water and moves, chiefly laterally, toward a point of discharge. Infiltration is produced by the joint action of molecular attraction and gravity. The rate of infiltration then becomes chiefly a function of the permeability of the soil. This permeability is almost infinitely diverse. Under conditions of unsaturated flow such as dominate in the zone of aeration, it varies with moisture content as well as with pore size. It varies also with the geology. For example, in the badlands of South Dakota, where the soil and rocks are of low permeability, the in-

the other hand, the soils of the Sand Hills of Nebraska and the glacial outwash deposits of Long Island are so permeable that they absorb the water of the most violent storms and permit little or no direct runoff.

The permeability of the rock materials beneath the soil zone also is important. Since the soil is commonly formed by weathering of the underlying rock, the permeability of the rocks is generally comparable to that of the soil.

Water-table and artesian conditions. Water that moves downward through the soil and subsoil in excess of capillary requirements continues to move downward until it reaches a zone whose per-

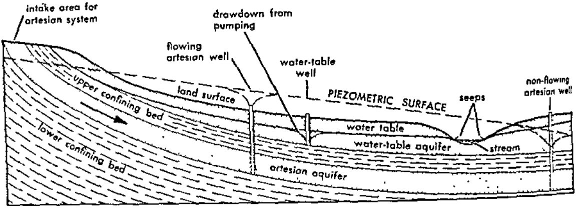


Diagram illustrating water-table and artesian conditions.

meability is so low that the rate of further downward movement is less than the rate of replenishment from above. A zone of saturation then forms, its thickness depending on the opportunity for lateral escape of water in relation to replenishment. The top of this zone is the water table. Under these circumstances, the water is said to be unconfined, or under water-table conditions. However, since much of the crust of the earth has a more or less well-defined layered structure in which zones of high and low permeability alternate, situations are common in which ground water moving laterally in a permeable rock passes between layers of relatively low permeability. Although the permeable layer contains unconfined ground water in the area where there is no impermeable layer (confining bed) above, the part of the layer, or aquifer, that passes beneath the confining bed contains water that is pressing upward against the confining bed, and if a well is drilled in this area the water in it will rise. It tends to rise to the level of the water in the unconfined area, but fails to reach that level by the amount of pressure head lost by friction as the water moves from the unconfined area to the well. Confined water is also called artesian water, and wells tapping it are called artesian whether or not their head is sufficient for them to flow at the land surface.

Chemical quality of ground water. Water is said to be the universal solvent. When it condenses and falls as rain or snow, it absorbs small amounts of mineral and organic substances from the air. After falling, it continues to dissolve some of the soil and rocks through which it passes. Thus, no ground water is chemically pure. Its commonest mineral constituents are the bicarbonates, chlorides, and sulfates of calcium, magnesium, sodium, and potassium, in ionized or dissociated form. Silica also is an important constituent. Common also are small, but significant, concentrations of iron, and manganese, fluoride, and nitrate. The concentration of the dissolved minerals varies widely with the kind of soil and rocks through which the water has passed. Ordinarily, water that contains more than 1000 parts per million (ppm) of dissolved solids is considered unfit for human consumption, and water that contains more than 2000 ppm, unfit for stock. However, both human beings and animals may become accustomed to greater concentrations. [A N S]

Bibliography: R. E. Horton, *The Field, Scope, and Status of the Science of Hydrology*, Trans. Am. Geophys. Union, 1931; O. E. Meinzer (ed.), *Hydrology*, 1942.

Ground-controlled-approach system (GCA)

A ground radar system providing information by which aircraft approaches may be directed via radio communication. A GCA system consists of a surveillance radar and an approach radar. See PRECISION APPROACH RADAR (PAR); SURVEILLANCE RADAR. [P.C.S.]

Grounding, electrical

A physical connection between non-current-carrying components and the earth. Most grounding schemes accomplish a threefold purpose.

1. They limit the voltage on metallic portions of electrical equipment to a safe value and thus eliminate a hazard to human life.

2. They provide a low-impedance return path to ground for currents resulting from certain types of faults within the equipment.

3. They prevent the accumulation of static charge on equipment, material, or personnel in areas where explosive hazards exist.

An ideal grounding system would consist of a single large conducting plate covering the entire area of interest and connected to the earth at an infinite number of points. This, of course, is impractical, but it can be effectively achieved with a suitable ground grid made of continuous conductors buried several feet below the surface of the ground and connected to driven ground rods. A continuous underground water piping system also provides a very satisfactory ground grid.

Two examples will better illustrate the need for equipment grounding.

An ungrounded motor in a factory in an area where the ground resistance is high develops a fault between one of its windings and its frame. A

high resistance between the fault point and ground will develop considerable voltage on the metal frame. A person coming in contact with the motor then becomes a parallel path to ground, and sufficient current may pass through his body to be fatal. With an adequate ground on the motor frame, a low resistance to ground will exist, thus limiting the voltage on the frame to a safe value.

Portable hand tools present a particular danger if not properly grounded, because the user is already holding the tool. If a fault develops in an ungrounded tool, the only possible path for the current to follow is through the user's body. For this reason, most hand power tools are now provided with special plugs that carry a ground and thus ensure positive grounding at all times. *See* CONNECTOR, ELECTRIC.

Accumulation of static charge presents a particularly serious problem in the chemical industry, where the movement of liquids and gases, and often the process itself, can cause charge to build up; the flour and grain industry, where an explosive mixture of dust and air could be set off by a small spark; the powder industry, where explosives must be protected; and in hospital operating rooms.

Grounding may not always be the solution to all these static problems, but many can be solved by bonding the various parts of the equipment together and then grounding the entire system. *See* GROUNDING, ELECTRONIC EQUIPMENT. For grounding of three-phase power systems, *see* ELECTRIC POWER SUBSTATION. [J.M.C.]

Bibliography: American Institute of Electrical Engineers, *Grounding of Industrial Power Systems*, AIEE Publ. 953, 1956; D. Beeman (ed.), *Industrial Power Systems Handbook*, 1955.

Grounding, electronic equipment

The provision of any low-impedance connection between electrical apparatus and the earth is known as grounding. The grounding may be satisfactory only for direct current or low frequencies, or it may be acceptable at relatively high frequencies *See* GROUNDING, ELECTRICAL.

Floating equipment. Apparatus which is not effectively grounded is said to float. A floating conducting body assumes the electric potential of the space surrounding the body. This potential is determined by the combined electric fields of the earth and those arising from other electric devices and wiring. The magnitude of the earth's electric field varies continually and erratically Electric fields of several thousand volts per meter are quite common during rainstorms; during electrical storms the electric field may rise to values on the order of 10,000 volts/m. Electric fields arising from other electrical equipment are usually periodic over short time intervals and of considerably lower field strengths.

Capacitance to ground. Every wire and electronic circuit component, as well as the chassis of an electronic device, possesses a finite capacitance

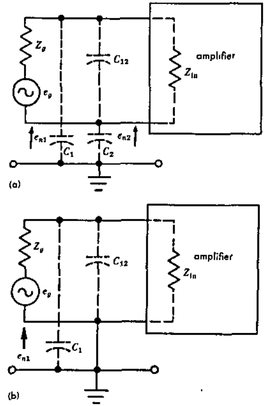


Fig. 1. An unshielded amplifier. (a) Floating (b) Grounded

to ground These ground capacitances have potentials appearing across them that are dependent upon the electrical noise fields present. The magnitudes of the capacitances are dependent upon the geometrical location of the device in relation to grounded objects, such as the earth, electrical wiring, buildings, furniture, and personnel. Relative motion between the device and any grounded object results in an erratically changing stray or ground capacitance.

A floating electronic device is shown in Fig. 1*a*, and Fig. 1*b* represents the same device with the common lead grounded. No shielding is assumed to be present. Z_g is the internal impedance of the signal source and Z_{in} is the input impedance of the amplifier. C_1 and C_2 are the input-lead-to-ground capacities while C_{12} is the interlead capacitance. Figure 3 represents the same amplifier housed within a conducting shield or chassis. C_g is the capacitance to ground of the shield or chassis, and C_1 and C_2 are now the lead-to-chassis capacitances.

Noise pick-up. Consider first the floating amplifier in Fig. 1*a*. The input to the amplifier is the normal signal voltage e_g plus a noise voltage e_n arising from noise voltages appearing across the stray capacitances. Noise voltage generated across C_{12} is relatively small and is neglected.

$$e_n = e_{n1} - e_{n2}$$

where the noise voltages developed across C_1 and

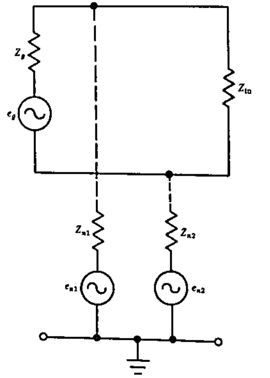


Fig 2. Approximate equivalent circuit of an unshielded, floating amplifier.

C_2 are probably of the same waveform and differ only in magnitude.

… for the float-
voltage input

$$e_{in} = \frac{e_g + e_n Z_g/(Z_{n1} + Z_{n2})}{Z_g/(Z_{n1} + Z_{n2}) + Z_g/Z_{in} + 1}$$

where Z_{n1} and Z_{n2} are the equivalent noise-source impedances. If the noise voltage e_n is to contribute only a negligible amount to the input, then the generator impedance must be small compared to the equivalent noise impedances. Further, grounding the common lead will set e_{n2} and Z_{n2} equal to zero but may only result in increased noise unless $Z_g \ll Z_{n1}$. See AMPLIFIER.

Shielding. If a conducting shield is placed around the amplifier, the shield becomes the local ground for the stray capacities C_1 and C_2 (see Fig. 3). The presence of the shield causes C_1 and C_2 to increase and, because C_2 is usually shorted out by a direct connection between the shield and the common lead, C_1 appears in parallel with C_{12}. This increases the total shunt capacitance across the circuit, thus lowering the bandwidth of the device (see VOLTAGE AMPLIFIER). A complete shield will eliminate the introduction of external noise signals e_{n1} and e_{n2}.

The shield is usually grounded to the earth or to a …

… example, must be provided with a glass window which represents a break in the shield. The stray capacitance to the ground plane through these breaks in the shielding can couple noise signals into the amplifier. Grounding the shield to the ground plane stabilizes the capacitive coupling through the breaks in the shield, so that relative motion of personnel and other grounded objects does not introduce noise voltages.

Second, bonding of any metallic object to ground prevents static voltages from appearing on the case. Such static voltages can produce unpleasant or even dangerous shocks to personnel. Equally important is the danger of sparking to other objects in explosive atmospheres.

Third, most electronic equipment derives its source of power from mains which are provided with ground connections. For instance, one side of the 120-volt, single-phase, ac main is usually grounded. If an insulation failure within the apparatus causes the shield to be placed in contact with the ungrounded power lead, a short circuit occurs and, if the shield is grounded, properly placed fuses will disconnect the apparatus. If the shield is not grounded, the shield takes the potential of the ungrounded lead and is a danger to personnel.

An exception to the above discussion is an electronic device employing an ac-dc power supply. These transformerless power supplies are employed in many radios, record players, and television sets. In this type of power supply one of the supply main leads is connected to the chassis of the apparatus. Depending upon how the user happens to plug the electrical socket into the wall outlet, the chassis is either grounded or at 120 volts ac above ground. Apparatus employing such power supplies therefore must be provided with a completely insulated chassis to protect the user. Signal grounding is accomplished by capacitive grounding, at audio or radio frequencies, of the chassis to the power leads. See SHIELDING, ELECTRICAL.

Ground-lead impedance. The impedance of each ground lead must be kept as low as possible, and particular care must be exercised to ensure that large currents, such as filament current, do not flow through the same ground impedance as do signal

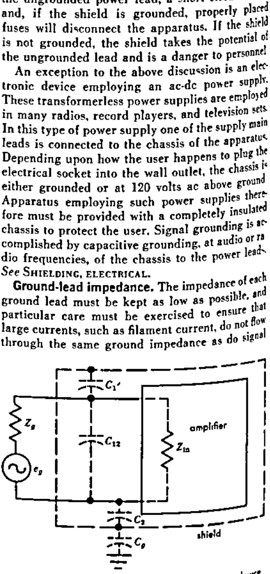


Fig. 3. A shielded amplifier or other electronic device

currents. For example, Fig. 4 depicts the coupling possible between a grounded heater lead and a grid lead by means of the extremely small impedance of a metal chassis.

The cathode and one heater lead are assumed to be bonded to the chassis at point 1. One side of the filament transformer is bonded to the chassis at point 3; therefore heater current flows through the chassis as shown by the current flow lines. Normal to the current flow lines are equipotential lines, as shown. The grid resistor is bonded to the chassis at point 2. A noise signal e_{12} is inserted into the grid circuit, resulting from the potential drop along the chassis between the cathode and grid bonding points. This drop may amount to only a few microvolts, but if this is one of the early stages of a high-gain amplifier then excessive noise may be noticed in the output.

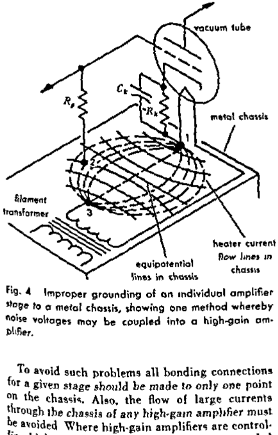


Fig. 4 Improper grounding of an individual amplifier stage to a metal chassis, showing one method whereby noise voltages may be coupled into a high-gain amplifier.

To avoid such problems all bonding connections for a given stage should be made to only one point on the chassis. Also, the flow of large currents through the chassis of any high-gain amplifier must be avoided. Where high-gain amplifiers are controlling high-power ac or dc circuits having grounded terminals, it is best to insulate the amplifier chassis from the frames of the rest of the machinery and then provide a single ground strap from chassis to frame. This will prevent large currents from passing through the amplifier chassis. *See* CIRCUIT, ELECTRONIC.

In general, each stage of a high-gain amplifier should have its own chassis ground. Each individual amplifier should have its own ground lead to a common amplifier ground point. All machinery and other power devices should be bonded together, with a single lead to the amplifier ground point. Finally, if additional shielding is provided, such as a metal cabinet, it should have a single ground lead to the common ground point for the amplifier and the power equipment. This common ground point would then be grounded to earth by bonding to a cold-water pipe or by other methods in conformance with the local electrical codes. [R.L.R.]

Ground-position indicator (GPI)

An electromechanical dead reckoning computer. Dead reckoning (a term derived from deduced calculation) is the procedure of advancing a known position by the addition of one or more vectors representing courses and distances. *See* DEAD RECKONING.

The ground-position indicator obtains directional information from an accurate compass which usually consists of a magnetic compass and a directional gyroscope. This device is located at some remote point, and the directional indication is carried to the cockpit by electrical means. Through the combination of the gyro and magnetic compass, the normal gyro drifts and the effects of turning, which cause large errors in the magnetic compass, are eliminated. Distances are obtained by integrating the instantaneous products of true air speed and time. True air speed is obtained by determining indicated air speed with a pitot tube and correcting for static pressure and air temperature. This correction may be made through electronic computer techniques; however, one mechanism accomplishes it by a variable-speed, motor-driven blower. The speed of the motor is automatically adjusted so that the pressure generated just counterbalances the pressure from the pitot tube. The speed of the motor is therefore proportional to the air speed.

A computer which advances a known position by a vector defined by the direction and air speed would give true position only if the air through which the aircraft travels were not in motion. Therefore, it is known as an air-position indicator. To obtain ground position, it is necessary to take into account wind direction and velocity. Wind direction and velocity can be manually inserted into the GPI. These quantities are determined either by the use of meteorological data or by flying a circular course and comparing the reading of the GPI with the correct reading when again over the known position.

Some GPI equipments are designed so that they accept input data from polar-coordinate or Doppler navigation systems. The data from these systems automatically introduce corrections for wind velocity and direction. GP indicators usually read position as latitude and longitude. Some of them include means for setting in the position of the desired destination. The equipment then reads course and distance to the desired terminal. *See* AIRCRAFT COMPASS SYSTEM; AIR-VELOCITY MEASUREMENT; NAVIGATION SYSTEMS, ELECTRONIC. [P.C.

Group theory

Any set of elements which is equipped with an operation (called multiplication) satisfying the requirements (1), (2), and (3) is called a group. Group theory is the branch of mathematics devoted to the properties of groups. Group theory has applications in the theories of relativity, quantum mechanics, and crystallography, and also in some branches of algebra and in analytic function theory.

Requirements on the group operation. The group operation is supposed to give for every pair of elements, for example, g and h, in the group, another element gh (called their product). Multiplication is supposed to satisfy the following requirements.

Associative law. If g_1, g_2, and g_3 are any elements of the group, then

$$(g_1g_2)g_3 = g_1(g_2g_3) \tag{1}$$

that is, in forming a product of three elements one obtains the same result by first multiplying g_1 and g_2 and then multiplying the result and g_3, as first multiplying g_2 and g_3 and then multiplying g_1 and the result

Existence of an identity element. There is in the group an element e (called the identity) with the property

$$eg = ge = g \tag{2}$$

for every element g of the group.

Existence of inverses For each element, g, of the group there is an element g^{-1} (called the inverse of g) which satisfies

$$g^{-1}g = gg^{-1} = e \tag{3}$$

Examples. An example of a group is the set of positive real numbers equipped with the operation of ordinary multiplication. The identity element is then the number 1 and the inverse of an element is its reciprocal. Another example of a group is the set of all real numbers equipped with the operation of ordinary addition. The identity element is then the number 0, and the inverse of an element is its negative.

Both these examples are commutative groups (also called abelian groups after the mathematician Niels Abel) because their multiplication law satisfies

$$gh = hg \tag{4}$$

for every g and h in the group. An example of a noncommutative group (also called nonabelian) is the group of rotations of a three-dimensional rigid body around a point. Here an element of the group is a rotation, that is, the act of rotating the body around a certain axis in space by a certain angle. The group operation means merely successive transformation: g_1g_2 stands for the act of rotation obtained by carrying out the acts of rotation g_2 and g_1 successively in that order. The identity element is the act of rotating through zero angle or making no rotation. The inverse of an element is the rotation around the same axis by the same angle but in the opposite sense. It is easy to check that if g_1 is a rotation around the vertical by 90° and g_2 a rotation around some horizontal axis by 90°, that $g_1g_2 \neq g_2g_1$, so the group is noncommutative.

These three examples are all of groups with an infinite number of elements, that is, infinite groups. There are also finite groups. A simple example is the group with two elements, the identity e and another e_1 satisfying $e_1e_1 = e$.

Topological groups. For infinite groups, it often occurs that a group has a natural geometry. For example, in the group of positive real numbers described above one has the geometry of the real numbers. This geometry is compatible with the group operations in the sense that the product gh is a continuous function of g and h, and the inverse g^{-1} is a continuous function of g. Generalizing this scheme to arbitrary groups one arrives at the notion of a topological group (or less frequently recently a continuous group) which is a set of elements equipped not only with a group operation but also with a topology, the two notions being required to be compatible in the above sense. A topology can be specified in a number of ways; the net effect of any of them is to give meaning to the phrase: two group elements are close to one another. For example, in the case of the rotation group discussed above one can define a topology by specifying that two rotations are close to one another when their axes have nearly the same direction and their angles of rotation differ by very little. *See* TOPOLOGY.

There is a particular class of topological groups which is of great importance. These are the so-called Lie groups (named after the mathematician Sophus Lie). These are topological groups in which it is possible to label the group elements by a finite number of coordinates in such a way that the coordinates of gh and g^{-1} are analytic functions of the coordinates of g and h and of g respectively. That means that they should be representable as convergent power series in those coordinates. Most of the topological groups occurring in applications are Lie groups. One important example of a group which is not a Lie group (because its elements cannot even be labeled by a finite number of coordinates in such a way as to make the group operation continuous) is the group of all permissible coordinate transformations in the general theory of relativity (*see* RELATIVITY).

Applications to physical sciences. The principal applications of group theory outside mathematics itself have to do with the classification and exploitation of symmetries in physical systems.

Crystallography. An important case is the problem of classifying crystal structures. To each crystal one associates two groups: its point group and its space group. The point group is the set of all those rotations of three-dimensional space with center in a fixed atom of the crystal, which carry atoms into identical atoms. The space group is the

set of motions of three-dimensional space which carry atoms of the crystal into identical atoms. The classification of the different possible symmetries of crystals reduces to the problem of finding all the different point and space groups. When the point and space groups of a crystal are known, one can say a good deal about the physical properties of the crystal, because the symmetry imposes restrictions on the possible form of such quantities as the elastic and dielectric constants. See CRYSTALLOGRAPHY.

Quantum mechanics. The classification of crystal structure can be carried out within the framework of classical physics (as well as that of quantum theory), but it is in quantum mechanics that the most extensive and significant applications of group theory exist. In quantum mechanics, the state of a physical system is described by a wave function, and a symmetry transformation (for example, a rotation or translation) of the state yields a new wave function related to the old by a linear transformation. For the particular case of a stationary state of a system, symmetric under the symmetry transformations considered, the new wave function must also be a stationary state of the same energy as the old: a symmetry transformation then carries wave functions belonging to a given eigenvalue of the energy into linear combinations of themselves. In any case, one is led to the problem of classifying the different possible transformation laws of a set of states under the given symmetry group. In mathematical terminology, the problem turns out to be that of classifying all representations of the symmetry group, that is, of finding all correspondences between group elements g and linear operators $M(g)$ such that

$$M(g_1)M(g_2) = M(g_1g_2) \qquad (5)$$

When the symmetry group is the rotation group, the classification of the representations can be made in terms of angular momentum; for the translation group in space, it is in terms of linear momentum. The classification leads to characteristic orthogonality relations and selection rules, the best known of which are the conservation laws of angular momentum and linear momentum. Another consequence is the theorem which says that a particle of spin s can have electromagnetic moments of order at most $2s$. For practical calculations of approximate wave functions of atomic nuclei, the techniques afforded by the theory of the representations of the rotation group have become nearly indispensable.

When the symmetry group of the physical system is the inhomogeneous Lorentz group, as is the case when the theory in question satisfies the requirements of the special theory of relativity, the classification of the representations is by parameters whose physical meanings are mass and spin. It leads to a classification of the possible wave equations for elementary particles. Although this theory is one of the best grounded and most satisfactory parts of existing elementary particle theory, the representations of the inhomogeneous Lorentz group, unlike those of the rotation group, have as yet been of little use in practical calculations. See SYMMETRY LAWS (PHYSICS).

Applications within mathematics. The applications of group theory within mathematics itself are numerous and important. The part of mathematics in which the notion of group was first clearly isolated was the theory of algebraic equations.

Algebra. Here, the basic idea is to associate a group (the so-called Galois group, named after the mathematician E. Galois) with the algebraic equation

$$a_0 + a_1x + \cdot \cdot \cdot + a_nx^n = 0$$

in such a way that the structure of the group reflects the type of operations necessary to compute the roots from the coefficients $a_0, \ . \ . \ . \ , a_n$. For example, using this method one obtains a simple proof that for $n > 4$, no solution in roots of the general equation exists.

Topology. In topology, groups are used to classify the structure of various geometric objects. The homology and homotopy groups serve this purpose for a manifold. As an example consider the homotopy group. To define it, one chooses a point P of the manifold and considers all continuous closed paths starting and ending at P. A notion of product of such paths is defined: if x_1 and x_2 are two paths, the product x_1x_2 is the closed path which is obtained by tracing out x_2 and then x_1. The inverse of a path is then the same curve traversed in the opposite direction and the identity path consists of the point P alone. Paths are divided into equivalence classes, two paths lying in the same equivalence class if one can be deformed continuously into the other. The notion of product then extends to equivalence classes, the product of two equivalence classes being obtained as the equivalence class of the product of any two representative paths. With this definition of product, the set of equivalence classes of continuous closed paths starting at P is a group, the (one-dimensional) homotopy group (also called fundamental group). Provided the manifold under discussion is connected (that is, provided any two points of the manifold can be connected by a continuous curve lying in the manifold), the fundamental group is essentially the same, independent of which point is taken as starting point for the paths. If the fundamental group of the manifold consists of one element only, the manifold is said to be simply connected.

Analytic functions. Group theory plays a funda-

eral complex variables, which is less understood. Only the former will be discussed here. The starting point is the fact that each analytic function has a natural domain of definition, its Riemann surface. There is another Riemann surface associated with it, its universal covering surface, which is sim-

ply connected. The universal covering surface may have several sheets for each sheet of its underlying Riemann surface, each point of the latter being replaced by several "lying above" it. Any function which is analytic on the underlying Riemann surface can be regarded as analytic on the universal covering surface; it will take the same values at those points which lie above a given point. The analytic function regarded as defined on the universal covering surface is invariant under transformations which carry the points lying above any point of the Riemann surface into themselves. These transformations form a group. Now a second basic fact is that any simply connected Riemann surface can be mapped in a one-to-one and analytic fashion onto one of three regions: the interior of the unit circle, the complex plane, or the complex plane closed by adding a point at infinity. Thus, an analytic function on an arbitrary Riemann surface can be regarded as an analytic function defined in one of these three regions and invariant under a certain group of transformations. Such functions are called automorphic functions. *See* COMPLEX NUMBERS AND COMPLEX VARIABLES.

Structure of groups. If G is a group and a subset G' of its elements is a group under the same law of multiplication, then G' is called a subgroup of G. Especially significant are the invariant subgroups which have the property that if g is any element of G' and h any element of G then hgh^{-1} is an element of G'. Let G and H be two groups and suppose that f is a mapping of G into H (that is, a function which for each element, g, of G yields an element $f(g)$ in H) with the property $f(gh) = f(g)f(h)$. Then f is called a [illegible]

[illegible] identity in H form an invariant subgroup, the kernel of the homomorphism f. If $H = G$ so that f [illegible]

[illegible] isomorphism gives precise meaning to the vague notion that two groups have the same structure. Evidently, a representation of a group, defined in Eq. (5), is a homomorphism of the group into the group of linear transformations of a vector space.

For topological groups it is natural to require of a homomorphism that it preserve not only the group structure but also the topological structure; for a topological group, a homomorphism is defined as a continuous mapping, f, of G into H satisfying $f(gh) = f(g)f(h)$. The group structure and the topological structure of a topological group impose severe restrictions on each other. Evidence for this statement is the solution of Hilbert's Fifth Problem (posed in 1900 and solved about 50 years later) which may be stated roughly: every topological group whose elements can be labeled by a finite number of coordinates so that the group operations are continuous is isomorphic to a Lie group.

For some topological groups it is possible to define a notion of integration on the group which has invariance properties under group multiplication. Namely, if F is a complex-valued function defined on the group and $\int F(g)\, d\mu(g)$ is the integral of F over the group with respect to a measure $d\mu(g)$, then $d\mu(g)$ is said to be left-invariant if

$$\int F(g)\, d\mu(g) = \int F(hg)\, d\mu(g)$$

holds for all integrable F and all h in the group. Analogously, right-invariance is defined by

$$\int F(g)\, d\mu(g) = \int F(gh)\, d\mu(g)$$

When a left-invariant measure exists so does a right-invariant and they are unique up to a multiplicative factor. For a compact group, that is, one for which $\int d\mu(g) < \infty$, left- and right-invariant measures coincide.

Invariant integration is a powerful tool in the study of representations of groups. Consider, for example, two irreducible representations of a compact group, G, by unitary matrices. (A representation is irreducible if there is no proper subspace of vectors carried into itself by all matrices of the representation. It is unitary if all the matrices are unitary.) Then

$$\int M^{(1)}(g)_{\kappa\lambda}\overline{M^{(2)}(g)}_{\mu\nu}\, d\mu(g) = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \frac{\delta_{\kappa\mu}\,\delta_{\lambda\nu} \int d\mu(g)}{\dim M^{(1)}(g)} \qquad (6)$$

where the first alternative holds if $M^{(1)} = M^{(2)}$ and the second if the γ are not equivalent ($M^{(1)}$ and $M^{(2)}$ are equivalent if there exists a unitary operator U such that $M^{(1)}(\gamma) = UM^{(2)}(\gamma)U^{-1}$ for all g in G); $\dim M^{(1)}(g)$ stands for the number of rows (or columns) in $M^{(1)}(g)$. The angular momentum selection rules referred to above are consequences of Eq. (6) in the case that G is the three-dimensional rotation group.

For Lie groups there is an important method of analysis, the so-called infinitesimal method. Consider, for example, a Lie group of matrices with matrix multiplication as group multiplication (Not all Lie groups are isomorphic to such groups, but many are.) A one-parameter subgroup is a subset of elements, $g(t)$, labeled continuously by a real parameter, t, and satisfying

$$g(t_1)g(t_2) = g(t_1 + t_2)$$

for all real t_1 and t_2. Such a subgroup can be written in the form $g(t) = \exp tx$. The matrix x is called the infinitesimal element of the one-parameter subgroup. If the group elements are labeled by n parameters one can choose n distinct one-parameter subgroups with linearly independent infinitesimal elements $x_1, \ldots, x_n$. These matrices will then satisfy commutation relations

$$x_i x_k - x_k x_i = \sum_{l=1}^{n} C_{ik}{}^l x_l$$

The constants $C_{ik}{}^l$ are called the structure constants of the Lie group, and they largely determine the structure of the group. *See* GRAPH THEORY; RING THEORY; SET THEORY.

[A S W]

Bibliography: G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, rev. ed. 1953; E. F.

Wigner, *Group Theory and Its Application to Quantum Mechanics of Atomic Spectra*, 1959.

Group velocity

The velocity of energy flow in a propagating wave, having dimensions of distance per unit time. Group velocity v_g can be expressed in terms of the phase velocity v_p by the equation

$$v_g = v_p - \lambda \frac{dv_p}{d\lambda}$$

where λ is the wavelength.

Group velocity is different from phase velocity in dispersive media where the velocity is dependent upon the frequency of the wave. For example, for waves at the surface of deep water, where the phase velocity is proportional to the square root of wavelength, the group velocity is equal to one-half the phase velocity. In the case of electromagnetic waves in dielectric media, the phase velocity also depends upon wavelength.

For sound waves in a gas, the group velocity is, for practical purposes, equal to the phase velocity under most circumstances. In a moving fluid medium, however, the group velocity is the vector sum of the local velocity of sound and the medium velocity, as compared with the phase velocity, which is given by the sum of the local speed of sound and the component of the medium velocity normal to the wavefront. *See* PHASE VELOCITY; WAVE MOTION. [W.J.C.]

Grouse

Any member of the family Tetraonidae, order Galliformes. This is a Holarctic family of 18 species, including some of the most important game birds. There are 9 species in the United States. Grouse are similar to quail, but differ in several details. The most obvious difference is the feathered tarsus of the grouse. Many grouse engage in drumming, dancing, and strutting during the mating season. The prairie chickens and ptarmigans are grouse, as are the several species, such as ruffed grouse (see illustration), blue grouse, and spruce grouse, with which the name is usually associated. *See* GALLIFORMES; PRAIRIE CHICKEN; PTARMIGAN; QUAIL. [J.D.B.]

The ruffed grouse, *Bonasa umbellus*; length to 19 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Grout

A mixture of cement and water called neat cement grout, or a mixture of cement, water, and fine aggregate that hardens with hydration of the cement to form a solid. The latter mixture is also called mortar. Grout may be pumped through drilled holes under pressure to fill voids or cracks in foundations for structures in order to improve load-bearing properties or watertightness of the material. Grout serves as a filling for joints in dams and for cable ducts in post-tensioned concrete structures. It may be used to solidify gravel or stone fill as in Prepakt concrete Grout containing fine aggregate is used to fill the small space between concrete foundations and the base plates of heavy machinery to provide uniform distribution of loading. Admixtures may be used to impart special properties such as pumpability. *See* CONCRETE. [I.L.T.]

Growing season

Generally considered to be the period of the year when climatic conditions are favorable for plant growth, common to a place or an area. In middle latitudes, with sufficient moisture, the growing season generally falls between the date of the last freeze in spring and the time of first freeze of autumn. This would apply to most annual plants and crops and to many perennials and biennials. However, some more hardy plants extend their growing season beyond these limits and continue some growth until the soil freezes; but some less hardy plants require minimum temperatures of 40°F as in the tropics or subtropics.

At the opposite temperature extreme there are areas where high temperatures during the summer (usually also accompanied by lack of rain) retard or prevent plant growth. Such areas commonly have mild winters so that the growing season may extend from the beginning of the autumn rains through the winter to the next dry season.

Some desert areas lack a dependable or well-defined growing season. During occasional but irregular rainy periods plant seeds burst into vigorous growth, covering the desert with a blanket of fast growing and flowering plants. These quickly produce seeds which then remain dormant until the next rain. Such growing seasons may be only a few days or weeks long and may occur only once in an unpredictable two, three, or several years.

These, however, are special cases. In most usages, the term growing season is of interest in connection with the principal annual food, fiber, and forest crops and, in common climatological practice, is the frost-free season. While this varies from year to year, it is possible to establish an

average date for beginning and ending and an average length for any location or area. Such averages are useful in comparing one location with another in terms of agricultural potential or in deciding the crop varieties suitable for an area. For example, cotton requires a growing season of from 180–200 days while many corn varieties require only 130–150 days.

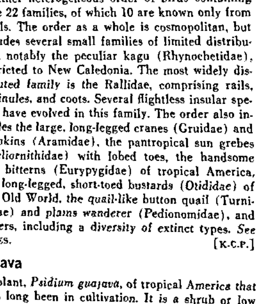
In the United States the average length of the growing season ranges from less than 100 days in the higher elevations of the Rockies to practically the whole year in southern Florida. As a rule, the length averages near 300 days along the Gulf and South Atlantic coasts and 120 days along the northern border. This is a decrease of about 10 days for every degree of latitude. [M.L.B.]

Gruiformes

A rather heterogeneous order of birds containing some 22 families, of which 10 are known only from fossils. The order as a whole is cosmopolitan, but includes several small families of limited distribution, notably the peculiar kagu (Rhynochetidae), restricted to New Caledonia. The most widely distributed family is the Rallidae, comprising rails, gallinules, and coots. Several flightless insular species have evolved in this family. The order also includes the large, long-legged cranes (Gruidae) and limpkins (Aramidae), the pantropical sun grebes (Heliornithidae) with lobed toes, the handsome sun bitterns (Eurypygidae) of tropical America, the long-legged, short-toed bustards (Otididae) of the Old World, the quail-like button quail (Turnicidae) and plains wanderer (Pedionomidae), and others, including a diversity of extinct types. *See* AVES. [K.C.P.]

Guava

A plant, *Psidium guajava*, of tropical America that has long been in cultivation. It is a shrub or low tree which belongs to the myrtle family Myrtaceae. The fruit is a berry, yellow when ripe, and quite variable in size depending on variety and growing conditions, the average being about 2½ in. long. The guava is quite aromatic, sweet, and juicy. It is used mostly for jellies and preserves, but also as a fresh fruit. *See* MYRTALES. [P.D.S.]



Cattley guava. (L. H. Bailey, *The Standard Cyclopedia of Horticulture*, vol. 2, Macmillan, 1937)

Guayule

This plant, *Parthenium argentatum*, of the composite family (Compositae) is a source of rubber. It is common in Mexico, New Mexico, and Texas. The latex occurs as solid particles—true caoutchouc (pure rubber)—suspended in the cell sap. To remove this, the dried plant is macerated with weak lye and then treated with a solvent. The yield obtained varies up to 22% of dry weight. Because of depletion of the foreign rubber supply during World War II, guayule was planted on some 36,000 acres in the southwestern portion of the United States. Production of guayule rubber, however, proved to be too costly to be an economic competitor of the rubber from the tropical plant *Hevea*. *See* CAMPANULALES; *see also* RUBBER; RUBBER TREE. [P.D.S.]

Guidance systems

Apparatus for guiding the motions of a vehicle, usually remotely and automatically. Guidance systems are necessary in missiles and spacecraft of all kinds, and in drone (pilotless) aircraft. They may also be employed in piloted aircraft, ships, and submarines, to relieve or assist the human pilot.

In common usage the term guidance implies supervision of the navigation of a vehicle on its path from present location to desired future location. The companion term, flight control, is used to connote control of the vehicle, such as in altitude, heading, and speed. The equipment which performs the flight-control function is called an automatic pilot, or autopilot, and is an essential component of every guidance system. *See* AUTOPILOT.

In addition to the autopilot, a guidance system consists of a computer for accepting information and determining desired flight path, and a sensor for measuring the actual flight path of the vehicle. The general arrangement is depicted in Fig. 1. The system's command computer determines the path to be flown and instructs the autopilot accordingly. The vehicle's actual flight path is measured by the sensor to be sure the guidance commands are being carried out, and errors are corrected by the usual feedback technique. *See* CONTROL SYSTEMS.

The representative case of command guidance of a ground-to-air missile is depicted in Fig. 2. The computer has been instructed to effect collision of the missile with the target aircraft. The computer is fed continuous information about the target position from one tracking radar, from which it computes the desired missile path. The other radar serves as the sensor of the actual flight path of the missile. The error between the desired and actual paths is computed and corrective signals are radioed to the missile autopilot.

Drone (pilotless) aircraft are often guided by ground radar tracking and radio command to the autopilot, using equipment such as that of Fig. 2. Even piloted aircraft are sometimes guided by ground equipment for special missions, such as target interception by fighter aircraft or automatic

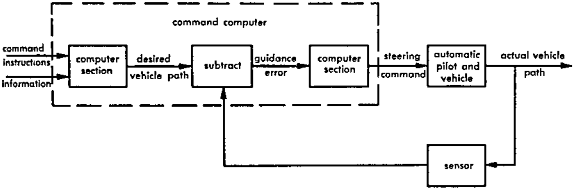


Fig. 1. Guidance system arrangement.

approach to a landing field. In the latter case either radar tracking, ground-controlled approach (GCA), or a system of fixed radio beams, instrument landing system (ILS), may be used to sense flight path. *See* FIRE-CONTROL SYSTEM; INSTRUMENT LANDING SYSTEM (ILS); PRECISION APPROACH RADAR (PAR).

In other short-range missile systems the missile may be caused to ride a radio beam to the target (beam-rider system); or a target-tracking radar or infrared seeker may be carried in the missile itself (homing system) or in a mother aircraft. *See* MISSILE GUIDANCE SYSTEMS.

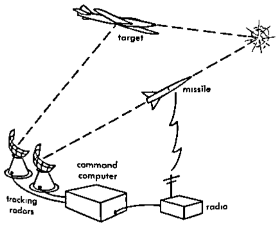


Fig. 2. Missile command guidance system.

In longer-range missiles inertial guidance equipment may be used. These employ accelerometers and gyroscopes in the missile to sense and keep track of its flight path (*see* INERTIAL GUIDANCE SYSTEM). Alternatively, automatic star-sighting telescopes, long-range radio navigation equipment, Doppler radar, or map-matching techniques may be used to perform the function of sensing the flight path of the missile (*see* NAVIGATION SYSTEMS, ELECTRONIC). In satellites and space vehicles an optical scanner may also be used to locate the edge of the earth or other planets. [R.H.C.]

Guided missile

A vehicle that carries automatic mechanisms which produce an effect upon the path it traverses. Such mechanisms may comprise all or part of the system provided for affecting the missile's path, but do not require control by a person aboard the missile. Generally, the missile is an unmanned vehicle, for destructive purpose, that travels above the earth's surface. It may, however, carry a passenger as its payload, be used for constructive, scientific purposes, or travel through a body of water as a torpedo (*see* MISSILE).

Guidance systems. The means for guidance of a missile generally consists of two intimately cooperating systems. The guidance system concerns itself with information about the path traveled by a reference point within the missile, such as its center of gravity. The control system deals with the velocity vector as determined by angular attitude and propulsion. The guidance (or navigation) system acts as the "brain and senses" which direct the control system in its capacity as the "muscle."

Guidance systems may be remote or self-contained. In remote systems, a portion of the guidance function is performed by equipment not carried by the missile. Self-contained systems perform all functions with on-board equipment.

Guidance systems are further classified as to whether they are preset or variable. In preset systems, data for a desired path or destination are inserted in the guidance mechanism prior to launching. Such a method is most practicable in directing a missile toward an immobile destination such as a fixed hostile missile launching site. Variable systems include means for changing the specified path or destination after launching. Such a change may be directed by remote or self-contained devices. Variable systems are most usefully employed against highly mobile targets, such as approaching enemy aircraft or missiles. *See* GUIDANCE SYSTEMS.

The guidance system has the primary duty of measurement of the missile position. Inherent to such a function is the need for reference coordinates oriented to serve as a basis for measurement.

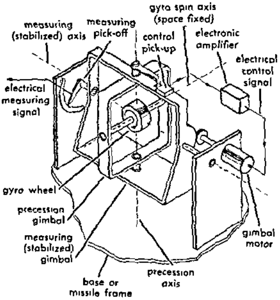


Fig 1. Gyroscopic stabilization for one axis.

of direction and distance. Such a reference system is almost always required to be three-dimensional. A *rectilinear coordinate* system, consisting of three mutually perpendicular axes, can be used with orientation fixed to the missile, to the earth, or to space. A spherical or polar coordinate system affords a means of measurement from a reference point by means of two mutually perpendicular angles and a radial distance. The reference point may be the missile launch position, its destination, or the instantaneous position of the missile itself (*see* NAVIGATION).

Inertial guidance. The dominant preset guidance system uses inertia; it is self-contained. The designation *inertial* guidance derives from the use of inertial devices that comply with Newton's laws of motion. These devices perform two functions. First, they establish a reference system, rectilinear and oriented to inertial space. Second, they sense acceleration.

Reference system. Gyroscopes mounted on a support that is free to rotate universally on three interconnected swiveling gimbals provide the reference. Such an arrangement, sometimes called a stabilized platform, provides for a measurement of the missile's angular attitude through use of the relative gimbal angles.

A typical arrangement for gyroscopic stabilization of a single gimbal (or measuring) axis is shown in Fig 1. In a universally stabilized platform, such an arrangement is provided for each gimbal axis. The gyro spin axis tends to hold the gimbal in alignment by the angular momentum of the spinning wheel. An external torque (such as that due to missile rotation), applied about the measuring axis, initiates angular precession about the precession axis. An electrical control pick-up senses the resultant rotation. The pick-up signal controls the gimbal motor, which applies a compensatory torque, nullifying the effect of the external disturbing torque.

Angular motion of the missile about the space-fixed gimbal is indicated by the measuring pick-off whose signal is applied to missile control mechanisms to effect the desired angular relation between missile and gimbal (*see* CONTROL SYSTEMS).

Acceleration. Inertial devices also sense accelerations of the missile. Any acceleration can then be successively integrated into velocity and distance traveled to solve the basic navigation problem.

The sensing of acceleration is performed by accelerometers mounted on the stabilized platform. Generally, three are used, oriented on the platform in accordance with the particular rectilinear reference system used. In a simple linear accelerometer, acceleration is measured by the force required to prevent the relative displacement of a test mass. A device more useful to missile guidance both senses accelerations and performs the first integration into a velocity indication. Such a device is the pendulous integrating gyro accelerometer. It is similar to the gyroscopic stabilizer system (Fig. 1) except that the base is mounted on a stabilized platform. Furthermore, a known unbalanced mass is mounted on the precession gimbal. In this case a linear acceleration (force) along the measuring axis produces a torque that induces the spin axis to precess about the measuring axis. Friction about that axis imparts a torque tending to cause an undesired precession about the precession axis. This undesired precession is compensated by the gimbal motor controlled by a signal from the pick-up. The cumulative angular rotation about the measuring axis is the time integral of acceleration and is proportional to velocity.

A second integration to furnish distance information can be performed by either a mechanical integrator or electrical circuits within the guidance computer. In the computer, velocity and distance information are compared to preset data, and the results are transmitted to missile control devices for action. The principal function involved is control of the direction of the velocity vector of the missile during the propulsion period, accomplished by positioning aerodynamic surfaces, positioning vanes within the exhaust jet, or swiveling of the rocket engine itself. The computer may also be used to determine the proper instant for motor cut-off and for arming of fuzing devices (*see* BALLISTIC MISSILE).

Use of inertial guidance as the sole means for navigation predominates in the long range vehicles such as the Jupiter, which is an intermediate-range ballistic missile (IRBM). An important advantage over other systems is its independence from external sources of transmitted intelligence susceptible to defensive jamming or interference. Inertial systems are subject to inaccuracies which generally are proportional to their operating time. These errors are random and unpredictable, except for their maxi-

mum possible value for a particular mission and operating time. They are caused by imperfect operation of the inertial devices and other parts of the system. The gyroscopic errors are caused by disturbances (usually frictional) that produce unwanted torques about the precession axis which result in drift of the spin axis away from its required direction in the established space coordinate reference system (*see* INERTIAL GUIDANCE SYSTEM). Some missiles use inertial guidance systems supplemented by external methods to correct periodically such accumulated errors.

Radio-inertial guidance. A preset remote form of guidance, such as is used in one version of the intercontinental ballistic missile (ICBM) Titan, combines inertial and radio techniques. The role of the inertial elements is secondary; they are used only for missile stabilization and to compensate for transient disturbances of flight path. Measurements of velocity and distance are made by radio or radar tracking from a ground station near the launch site. Ground-based computers determine errors in missile velocity and position and transmit corrective commands to receivers in the missile. These commands, blended with data from the inertial elements, operate the missile control devices (*see* RADAR).

Radio is used with guided missiles in various modes for several purposes. Telemetering is the transmission of internal missile measurements to ground stations. Command is the transmission of instructions from ground or other controlling station to the missile. Radar direction finding and range finding techniques use the direction and elapsed time for the echo from a missile or target of a previously transmitted signal. Another common technique includes the use of missile-borne beacons which reply to radio signals transmitted from ground bases.

An advantage of radio guidance systems is that the computer need not be carried by the missile and may consequently be elaborated for greater inherent accuracy. Disadvantages are the susceptibility to jamming and the deterioration of accuracy in tracking the missile as its range from the ground station increases (*see* MISSILE GUIDANCE SYSTEMS).

Stellar-inertial guidance. Errors that accumulate in a preset, self-contained inertial guidance system can be corrected by data observed by star trackers. These consist of several automatic telescopes oriented to the inertial space reference established by the stabilized platform. Photoelectric devices sense tracking errors, which are processed in the guidance computer into angular corrections for the appropriate stabilizing gyros. The angular orientation of the telescopes relative to the platform must be adjusted immediately prior to launch for sidereal motion (rotation of the earth with respect to the stars). The instant of launch initiates the use of the inertial space reference system, relative to which the star angles remain fixed. Automatic star trackers can track stars during broad daylight as well as during hours of darkness. Stellar-inertial guidance is used by the intercontinental range Snark, an air-breathing missile of the pilotless-aircraft type.

Other preset systems. Terrain recognition is a preset guidance system which uses map-matching. This technique is an automatic comparison of the radar image of the terrain under the missile with a map previously prepared on a negative transparent film. The radar image is projected through the film onto a photoelectric tube. Coincidence is determined when the transmitted light is a minimum. Left-right errors cause lateral movement of the film carriage, and errors parallel to the missile's course change the speed of translation of the film. These corrections are applied to the corresponding missile controls, and the errors are thus cancelled.

Hyperbolic guidance is a radio navigational system adapted from that used by surface craft and piloted aircraft. The difference in the time of reception of signals transmitted from two or more stations is measured. This corresponds to a difference in distance and, for a given pair of stations, establishes a line of position (in the form of a hyperbola) along which lies the missile's position. The intersection of two such lines of position establishes a fix, or the missile's geographic location (*see* LORAN).

Terrestrial field guidance systems are based on measurement of direction and intensity of the earth's magnetic and gravitational fields. Although of limited importance for military purposes because of lack of accuracy, they may prove helpful in navigation of satellite vehicles.

The guided missiles described above are inherently preset. Those to be discussed below are basically of the variable type. They are more generally apt to have different types of guidance used during the principal phases of their travel from launch to impact. These phases are launching, midcourse, and terminal (approach to target). However, mid-

as that required to prevent roll. This serves to prevent interchange of control signal coordinates, as would occur, for instance, if a missile were to roll onto its side during flight so that an externally generated guidance signal intended to tilt the missile up produced instead a right turn.

Command guidance. In a usual form of variable guidance, intelligence is generated remotely by sensing and computing devices at the launching station from which control signals are transmitted to the missile.

Radio command. In the method most used for surface-to-air missiles, two tracking radars follow the target aircraft and the missile (Fig. 2). The measured angles (azimuth and elevation) and the ranges are fed into a computer. It computes the control orders which will enable the missile to hit

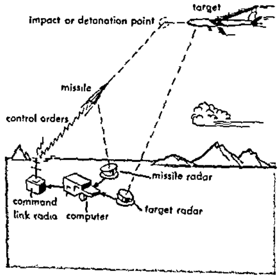


Fig. 2. Radio-command guidance system.

the target or pass close enough to it to make a kill using proximity fuzes. The orders are transmitted through a command link radio, a receiver in the missile, and then to the missile control mechanisms. Tracking, computing, and command functions are performed automatically with a minimum of monitoring attention by ground crews. The path computed for the missile is one of minimum curvature to avoid stresses from high lateral accelerations. This is done with a lead technique in which the missile follows a collision course. As in surface navigation, the relative bearing of the target from the missile remains constant. Radio command systems are sometimes used in simultaneous combination with other techniques, as in radio-inertial guidance, mentioned above. They are also used sequentially with other systems, particularly with respect to the terminal phase.

Wire command. A simple form of remote guidance, originally tried by the Germans for air-to-air weapons, uses wires to the missile. The technique is limited to short ranges. The missile is tracked relative to the target by optical or acoustic means. The control commands are transmitted on a connecting wire reeled out from the missile during its travel. This technique has been used with the Dart antitank missile and with antisubmarine torpedoes.

Beam rider. Like radio command, beam rider guidance uses a radar to track the target. However, in this case, the missile follows the radar beam to the impact point.

The missile carries electronic equipment capable of determining its position relative to the center of the beam. The beam traces a circle about the reference scan axis. When the missile is centered on the scan axis, it receives a signal of constant amplitude. When it is not centered, it receives an error signal, modulated in phase with the scanning motion. The error signal is compared with a reference amplitude-modulated signal of the same frequency as that of the scan to return the missile to the scan axis.

Beam-rider guidance has been used for air-to-air missiles, but imposes a disadvantageous tactical restriction on the launching aircraft, which must continue the control until impact. Its use is more general in surface-to-air missiles, such as the Navy's antiaircraft Terrier.

Homing guidance. Those systems in which the missile carries equipment for sensing radiation emanating or reflecting from the target are homing guidance systems. Radio waves, heat, sound, or light may be sensed. Three modes are used. The active mode is that using waves transmitted from the missile and reflected back to it by the target. The semiactive mode uses reflected waves, originally transmitted by a source other than the missile (Fig. 3). The passive mode uses waves originating at the target. In the active and passive modes, the system is self-contained in the missile. Homing guidance is of particular advantage to the terminal phase because the accuracy increases as the missile nears its target. Homing guidance is often used after other modes of guidance have brought the missile near the target.

Radar homing. By sensing electromagnetic energy, radar homing employs all three modes in guidance of missiles with various missions. However, the passive mode, which requires transmission from the target, has few applications; one is attack from the air against naval targets.

Infrared homing. By sensing heat waves from a target, infrared homing enables a missile to discriminate a target from the environment in much the same manner as does optical guidance. The receiver uses heat-sensitive detectors such as the thermistor bolometer, whose electrical resistance is changed by the warming of the incoming radiation. It is a thin flake composed of mixtures of semiconducting metallic oxides of large temperature coefficient of resistance. The Sidewinder air-

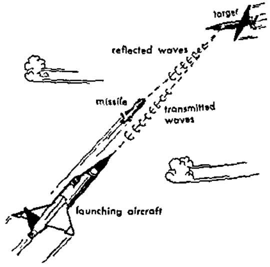


Fig. 3. Semiactive radar homing guidance system

to-air missile employs infrared homing guidance (*see* INFRARED DETECTOR).

Acoustic homing. Sound waves are used for guided missiles whose terminal phase occurs underwater. Torpedoes and antisubmarine missiles are of this category. *See* ACOUSTIC TORPEDO.

Optical homing. Guidance based on radiation of light waves finds few missile applications. However, TV homing (quasi-optical) has interesting possibilities. The missile-aiming error relative to the target is sensed pictorially by a television camera in the nose of the missile. The picture, transmitted to the launching aircraft or a ground station, is used to determine corrective steering orders sent to the missile.

Environment. A guided missile is exposed to severe environments, including extremes of acceleration, vibration, temperature, and altitude during its functional life as well as to shock, humidity, and exposure to sand and dust during military handling and storage. The nature of in-flight environmental conditions is usually determined during the developmental period by test firings of the propulsion system and the missile frame (*see* BALLISTIC RANGE; ROCKET SLED TESTING). Measurements are made at critical locations by transducers that transform the characteristic readings into electrical signals, which are telemetered to ground stations. Elements of the guidance system are subjected to laboratory tests simulating the flight environment (*see* ENVIRONMENTAL TEST). Design corrections are incorporated as needed to achieve proper operation under conditions in excess of those measured to provide a safety factor.

In subsequent test flights using guidance, measurements of data within that system are similarly made. These afford a means of analytical evaluation of system performance at the subsystem and component level. External tracking measurements of the missile's position and velocity are made from launch site and down-range stations to evaluate over-all missile performance.

Reliability. In a guided missile, reliability is more important than in a manned aircraft where use of alternate devices is possible. Durability is of less significance because of the shorter usage period.

Successful performance by a guided missile is possible only if it is highly reliable. This means more than simply completing its cycle of functions without an outright failure; this high quality of performance must result in the specified accuracy, which is almost entirely determined by the guidance system. In even the simplest system, the guidance equipment consists of several subsystems, each containing dozens of components comprised of hundreds of parts. The optimum reliability attainable by a guidance system is established by its design. It is essential that complexity and the use of marginally safe components be a minimum compatible with accuracy requirements and limitations of weight, volume, power consumption, and cost (*see* RELIABILITY OF EQUIPMENT). Extreme attention is required in the choice and application of components ranging from those of simple function (resistors, capacitors, potentiometers, relays, motors) to the higher order assemblies (gyros, accelerometers, transistor amplifiers, magnetic tape recorders, radar magnetrons, digital computers).

The level of reliability of a guidance system, established by system design and use of qualified components, can be realized in ultimate missile performance only if adequate quality control is applied to its fabrication (*see* QUALITY CONTROL). Of particular importance in electromechanical and electronic assemblies is the soundness of electrical connections at soldered joints and plug connectors. Use of mechanized assembly has considerably reduced the number of such connections, which are subjected to meticulous process control and inspection to avoid failure induced by missile vibration (*see* PRINTED CIRCUIT).

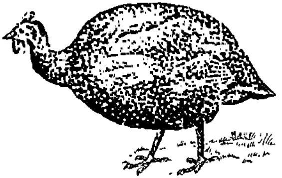
Check-out equipment. A guided missile is given a sequential series of operational tests and calibrations to assure that it is capable of performing all of its functions. This is done after final missile assembly, periodically during storage, and usually just before launching (*see* LAUNCHING PAD COMPLEX). The guidance and control systems are usually automatically cycled through a series of operations and measurements

Alignment of a guided missile (or, in particular, its guidance system) is of various degrees of importance, depending on the type of missile. The inertial guidance system of a ballistic missile must be precisely aligned vertically and in horizontal azimuth because of its mission to impact a target of known geographic location. Because the stable platform is the reference of orientation, it must be precisely aligned to earth coordinates at the instant of launching, after which it maintains a reference fixed to space. Vertical alignment can be maintained with vertical-seeking (pendulous) devices on the platform Azimuth alignment of the platform can be maintained, relative to presurveyed geographic angles, using an automatic theodolite. It observes the reflections from a mirror mounted on the stable platform. Observed rotational errors are converted into signals that are applied to drive elements of the platform to correct alignment. [L.S.B.]

Bibliography: A. S. Locke et al., *Guidance*, 1955; U.S. Department of the Air Force, *Guided Missiles Operations, Design and Theory*, 1958.

Guinea hen

A pheasant, *Numidia meleagris*, of the family Phasianidae, originally from the West African jungles. This bird, also known as the guinea fowl, has been kept as a domestic fowl since the Middle Ages. The common, or pearl variety, is slate gray and is marked with numerous white spots. There are also white and lavender varieties. Guineas are kept both for ornament and because some people enjoy hearing their unusual call. The birds are



The guinea fowl, *Numidia meleagris*. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

marketed in limited numbers, but are readily sold because of the similarity of their meat to that of wild game birds. *See* GALLIFORMES. [J.D.B.]

Guinea pig

A medium-sized, short-tailed rodent, of the family Caviidae, also called cavy. There are about 20 wild

as a laboratory animal in medical and biological experiments, and is also kept as a pet.

The guinea pig, *Cavia cobaya*; length to 10 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

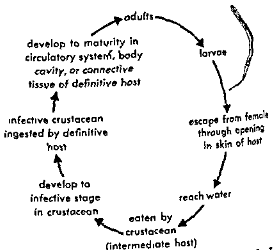
Guinea pigs have short ears and rudimentary tails. The fur of some domestic breeds is long and silky and the animals are highly variable in color. The wild pattern apparently was grayish brown and the fur rather coarse. Adults weigh about 1 lb, and are 9–11 in. long. *See* RODENTIA. [J.D.B.]

Guinea worm infection

An infection in man with the nematode *Dracunculus medinensis*; also called dracunulosis or dracontiasis. The male worms are only 1 in. long, but females grow up to 1 yd long. The females develop in 8–12 months in the body cavities and migrate finally to the subcutaneous tissues, preferably of the limbs. A blister with a central papule forms at a spot where the worm mouth approaches the epidermis. It bursts open, particularly upon contact with water, and numerous larvae are liberated. A small crustacean, the water flea (*Cyclops* and related genera), ingests the larvae and acts as an intermediate host. Man becomes parasitized upon drinking well water containing these crustaceans.

Allergic and toxic reactions such as erythema, generalized appearance of intensely itching welts, severe pruritis, vomiting, and diarrhea occur a few hours before the papule bursts. The disease is and has been associated for centuries with religious rites. Various animals act as reservoirs, even in America.

The disease is diagnosed by examining a drop of water which has been placed on the papule. The



Nematode cycle, guinea worm type. (From T. T. Mackie, G. W. Hunter, and C. B. Worth, *A Manual of Tropical Medicine*, 2d ed., Saunders, 1954)

presence of the larvae in the water is evidence of the infection.

The disease is treated with some success with phenothiazine injections. Another method of treatment requires rolling the worm out inch by inch on a small stick.

The disease is avoided by not drinking water containing infected crustaceans. The presence of *Cyclops* in wells is controlled by the use of quicklime, copper sulfate with perchloron, or minnows which feed on the crustaceans. *See* DRACUNCULOIDEA; PARASITOLOGY, MEDICAL. [J.F.MA.]

Gulf of Mexico

The Gulf of Mexico, a structural basin, has a surface area of 616,000 sq mi. The maximum depth, approximately 12,500 ft, is in the Sigsbee Deep. There is much conjecture on the origin of the Gulf. The basement rocks may be an extension of the continental mass. Following the Cretaceous Period this basement may have sunk, creating the present Gulf. Some studies indicate however, that the Mohorovicic discontinuity rises under the Gulf, implying basin permanency. *See* MOHO (MOHOROVICIC DISCONTINUITY).

Submarine structures and deposits. The coastlines are of three basic types. Florida and Yucatan appear as drowned limestone plateaus. The north

ern coast line is alluvial and the Mexican shore is of recent orogenic origin.

The most extensive continental shelves are off Texas, western Florida, and Yucatan. The width off the Rio Grande River is 45 miles, widening to 120 miles near the Texas-Louisiana border. The Florida and Yucatan shelves are about 120 and 135 miles wide, respectively (Fig. 1).

Fig. 1. Bathymetric features of the Gulf of Mexico (Adapted from U.S. Fish and Wildlife Serv., Gulf of Mexico, Fishery Bull. 89, 1954)

Numerous prominences dot the shelf and slope of the northwestern Gulf with reliefs ranging from 12 to a few thousand feet. Some are due to underlying salt domes. The structures on the slope appear to be due to slumping and faulting. The western Florida and Yucatan slopes are escarpments, originating by faulting or limestone cap-rock development.

Three principal submarine canyons dissect the terrace: the Mississippi, the DeSoto, and the Campeche canyons During the lowered sea level, channels were eroded into the offshore Pleistocene deposits of the northwestern Gulf and Texas bays. *See* FLOOD PLAINS, SUBMARINE CANYON.

In one month the Mississippi River may discharge 78.000,000 tons of sediments at Southwest Pass The principal clay is montmorillonite, which is common in the deposits, while most oceanic areas contain illite. The northwestern shelf consists of sand, silt, and clay. Slope sediments are similar but finer-grained The Florida and Yucatan shelf consists principally of detrital carbonates. Deep-water deposits are characterized by muds and

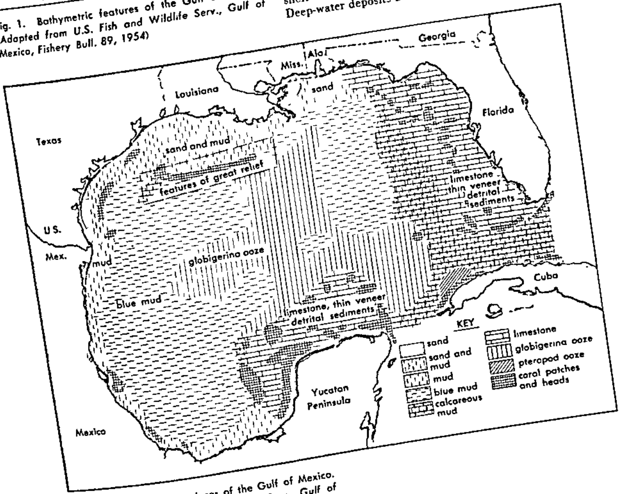


Fig. 2. Sedimentary provinces of the Gulf of Mexico. (Adapted from U.S. Fish and Wildlife Serv., Gulf of Mexico, Fishery Bull. 89, 1954)

calcareous oozes (Fig. 2). *See* DELTA; MARINE SEDIMENTS.

Temperature, salinity, and currents. Northern coastal surface water temperatures in midwinter approximate 65°F. The central and southern Gulf are about 73°F. The midsummer average is about 84°F.

The surface salinity of the central Gulf is approximately 36 00 parts per thousand (‰). However, the Mississippi region may have salinities of 26.0‰, 50 mi from shore. Higher salinities near western Florida and Campeche Bank may be explained by upwelling of Caribbean water passing through Yucatan Strait.

Some surface waters entering through Yucatan Strait spread across the Gulf, converging near the south Texas shore. The main stream moves toward the Mississippi, turning eastward and leaving through the Straits of Florida. *See* CARIBBEAN SEA; GULF STREAM.

The astronomical tide is usually 1–2 ft. Hurricanes and associated storm surges may cause local levels to exceed 15 ft while "northers" may lower the levels in bays by 4 or more ft. *See* STORM SURGE.
[R.G.B.]

Florida Straits. The Straits of Florida connect the Gulf of Mexico to the Atlantic Ocean. Both width and depth decrease northward (Fig. 3). The swift current in the Florida Straits, exceeding four knots, derives its energy from a downhill flow, though the Key West–Miami sea level drop does not reflect the downstream acceleration. It is the only permanent, swift ocean current in the world walled in on both sides.

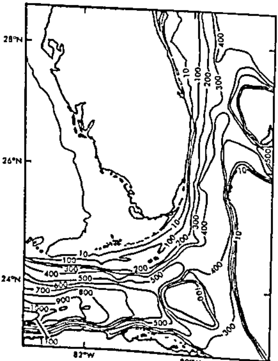


Fig 3. Straits of Florida, general bathymetry. Depths in fathoms. (From M. P. Wennekens, *Straits of Florida, Bull. Marine Sci. Gulf and Caribbean*, 9(1) 8,1959)

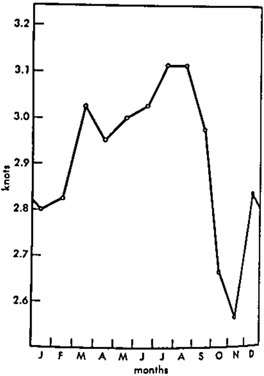


Fig 4. Annual variation in the average speed of the Florida Current between Miami and Gun Cay.

The first detailed current measurement in the Straits, made in 1885–1886, disclosed that the current axis is west of midstream, or the geostrophic axis of G. Wüst. This has been explained as involving the cross-stream component of the stream motion which was also noticed in 1885–1886 Two constituents of the cross-stream component have been found. One is related to the lateral wall friction and led to an inferred spiral path for the water. The other is identifiable with internal seiches of near-tidal periods and consequently complicates the geostrophic computation of the current, which yields an average transport of 26×10^6 m³/sec and a summer maximum (Fig. 4).

Some unexplained features are the significance of direct influx of West Atlantic water; the large transport fluctuation, 15–54 × 10^6 m³/sec, off Key West as reflected by the current's electric potential; and the frequent occurrence of biaxial cross-current profile. *See* OCEAN CURRENTS; SEICHE.
[F.C.]

Bibliography: U.S. Fish and Wildlife Serv., *Gulf of Mexico: Its Origin, Waters, and Marine Life*, Fishery Bull. 89, 1954.

Gulf Stream

A great ocean current transporting about 70,000,000 tons of warm water per second northward from the latitude of Florida to the Grand Banks off Newfoundland, an amount of water 1000 times

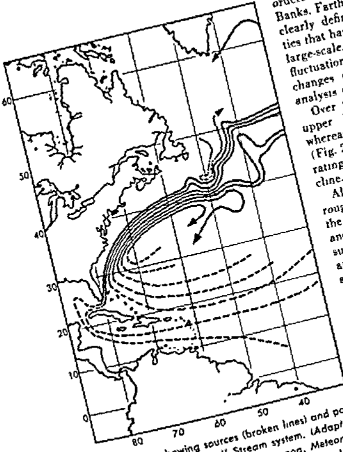


Fig. 1. Sketch showing sources (broken lines) and pattern (solid lines) of the Gulf Stream system. (Adapted from C. O'D. Iselin, Papers Phys. Oceanog. Meteorol., Mass. Inst. Technol. and Woods Hole Oceanog. Inst., 4(4):1–99, 1936)

the discharge of the Mississippi River. Before the days of scientific oceanography it was supposed that the origin of the water in the Gulf Stream was the Gulf of Mexico—as indeed the name implies; but now the origin has been definitely traced much farther upstream, through the Gulf of Mexico and Caribbean, to the Great North and South Equatorial currents of the Atlantic (Fig. 1). In fact, it seems much more reasonable to think of the Gulf Stream as a portion of a great horizontal circulation in the ocean, where each particle of water executes a closed circuit, sometimes moving slowly in midocean regions and other times rapidly in strong currents like the Gulf Stream. Thus there can scarcely be a meaning to the concept of the beginning and end of the Stream other than simple arbitrary geographical limits.

Description. The Stream is as deep as 610 m, is 20–40 miles wide, and moves at velocities up to 4 miles an hour. It does not stay in a fixed path but meanders like a river. The meanders generally are 100–200 miles long and move downstream at a rate of about 5 miles a day. The large-scale meanders first become noticeable after the Stream leaves the coast at Cape Hatteras and increase in amplitude until the pattern of flow has become extremely disordered and complex in the longitude of the Grand Banks. Farther to the east the field of motion is not clearly defined. There are small-scale irregularities that have been detected by aerial surveys; and large-scale, long-period irregularities and some fluctuations evidently connected with seasonal changes of the winds have been detected from analysis of tide gage records.

Over the tropical and subtropical Atlantic the upper layers of water are warm (10–27°C), whereas the deep waters are as cold as 2.2°C (Fig. 2). There is rather an abrupt transition separating these two layers, the so-called main thermocline. *See* THERMOCLINE.

Along the United States coast and along a roughly defined line between Cape Hatteras and the Faeroes the thermocline rises to the surface and north of it all the water is cold, even at the surface. The Gulf Stream flows along this boundary immediately on the warm-water southern side, and the Coriolis force acting upon the Gulf Stream (to the right of its direction of flow) balances the northward pressure gradients that otherwise would force the warm water above the thermocline to flow much farther northward. In this sense the Gulf Stream acts as a "dynamic dam" to prevent warm surface waters from further flow northward, so that an increase in Gulf Stream discharge might actually be construed as leading to cold Northern European climate rather than the reverse as is so often asserted in lay literature.

Theory. In other oceans there are also strong currents on the western sides, such as the Kuroshio in the North Pacific and the Agulhas in the South Indian. The marked tendency toward an east-west asymmetry in the horizontal circulations of all the oceans (an intensification of currents in the west)

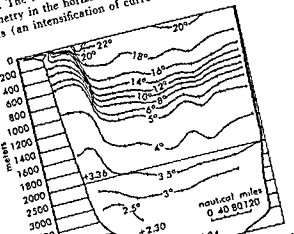


Fig. 2. Temperature (°C) section across the Gulf Stream, Chesapeake Bay to Bermuda, April 17–23, 1932. Vertical exaggeration X370 above 2000 m; X148 below 2000 m. (Adapted from C. O'D. Iselin, Papers Phys. Oceanog. Meteorol., Mass. Inst. Technol. and Woods Hole Oceanog. Inst., 4(4):1–99, 1936)

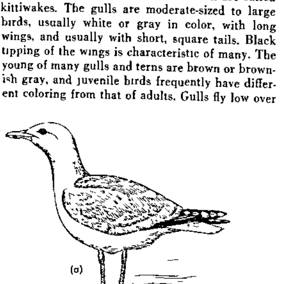
has been a subject of theoretical investigation in the decade 1948–1958. The theoretical explanation of broad features of the general oceanic circulation now seems to have been established. In the particular case of the Gulf Stream it appears that about half of the transport is due to the prevailing wind system over the North Atlantic, whereas the other half is tied in with a world-wide system of currents driven by thermal differences and is fed by surface waters from all over the globe. Associated with the thermal part is a deep cold countercurrent under the Gulf Stream, most of which crosses the Equator into the South Atlantic and thence is distributed into the cold deep waters of the Indian and Pacific Oceans. *See* ATLANTIC OCEAN; CARIBBEAN SEA; GULF OF MEXICO; OCEAN CURRENTS.

[H.SL.]

Bibliography: H. Stommel, *The Gulf Stream*, 1958.

Gull

Any of about 50 different species of the subfamily Larinae, family Laridae, a few of which are called kittiwakes. The gulls are moderate-sized to large birds, usually white or gray in color, with long wings, and usually with short, square tails. Black tipping of the wings is characteristic of many. The young of many gulls and terns are brown or brownish gray, and juvenile birds frequently have different coloring from that of adults. Gulls fly low over the water in search of fish near the surface which they catch with a sudden dive, much in the same manner as terns. Gulls rest and roost on the water, frequently in rather large flocks.



(a) Herring gull, Larus argentatus; length to 26 in. (b) Laughing gull, Larus atricilla; length to 17 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Gulls are world-wide in distribution and most of them are marine, although there are a few species found both along the seacoast and on inland waters. Franklin's gull, *Larus pipixcan*, is almost exclusively a fresh-water gull during the breeding season, and is quite common on the lakes of the northern prairies in the United States and Canada. The ring-billed gull, *L. delawarensis*, is another fresh-water summer resident, breeding all across northern North America from Alaska to the Great Lakes.

Best known of all gulls is the herring gull, *L. argentatus*, a Holarctic species, common on both fresh water and along the seacoast throughout the northern half of the United States. This is the gull that most commonly follows ships at sea. *See* CHARADRIIFORMES; TERN. [J.D.B.]

Gum

A class of high-molecular-weight molecules, usually with colloidal properties, which in an appropriate solvent or swelling agent are able to produce gels (highly viscous suspensions or solutions) at low dry-substance content. The molecules are either hydrophilic or hydrophobic; that is, they do or do not have an affinity for water. The term gum is applied to a wide variety of substances of gummy characteristics, and therefore cannot be precisely defined.

Various rubbers are considered to be gums, as are many synthetic polymers, high-molecular-weight hydrocarbons, or other petroleum products. Chicle for chewing gum is an example of a hydrophobic polymer which is termed a gum but is not frequently classified among the gums. Quite often listed among the gums are the hydrophobic resinous saps that often exude from plants and are commercially tapped in balsam (gum balsam) and other evergreen trees (gum resin). Incense gums such as myrrh and frankincense are likewise fragrant plant exudates.

Usually, however, the term gum, as technically employed in industry, refers to plant polysaccharides or their derivatives. These are dispersible in either cold or hot water to produce viscous mixtures. Modern usage of the term includes water-soluble derivatives of cellulose and derivatives and modifications of other polysaccharides which in the natural form are insoluble. Usage, therefore, also includes with gums the ill-defined group of plant slimes called mucilages. *See* CELLULOSE; COLLOID; POLYSACCHARIDE.

Viscosity. Gums are important in that they impart viscosity to aqueous solutions. Their physical properties are manifestations of their chemical structure, the kind and amount of solvent, and the kind and concentration of ions and other substances dissolved in the solvent. Because gums are

commonly composed of several different kinds of monomer units with many possible variations in regard to degree of branching, length of branches, and types of linkages, an almost infinite number of structures is possible. Forces act between molecules, between different parts of the same molecule, and between polymer and solvent. These forces include hydrogen bonding, ionic charges, dipole and induced dipole interactions, and van der Waals forces.

All of these forces affect such properties as gel-forming tendency, viscosity, and adhesiveness. The types of linkage due to their effects on chain flexibility are important also in determining physical properties. For example, it is known that linear molecules make more viscous solutions than do long-branch molecules of similar molecular weights, but they have a tendency to precipitate because of association of the chains. If this association is prevented, stability can be achieved without much sacrifice of viscosity. This can be done by introducing groups with ionic charges that repel one another, or by attaching many very short side branches to prevent close approach of the chains. Much work remains to be done in this area. While it is known that solutions of some gums are slimy or mucilaginous, whereas those of other gums are tacky, the reasons for these differences are unknown. Rheological properties of the different gum solutions also differ. *See* HYDROGEN BOND; POLARIZATION (DIELECTRICS); RHEOLOGY.

Source and structure. Many plant gums originally used by man still are important items of commerce. Most of these are dried exudates from trees and shrubs, produced with or without artificial stimulation by injury. These gums are collected by hand, usually in the hot, semi-arid regions of the world, often from wild plants, but sometimes from cultivated plants. For the most part, these gums have highly branched structures containing two to five different sugar units. They frequently are acidic in that they contain carboxyl groups which are portions of glycuronic acids. Commercial gums often are mixtures of two or more different polysaccharides.

Seed gums, which also have been used for many centuries, likewise are important items of commerce today. The more ancient gums were extracted from quince, psyllium, or flax seeds. The ground endosperm of locust trees grown in the Mediterranean area also has been an item of commerce for many years. It is still important today, along with a similar gum obtained from the endosperm of the guar plant which is native to India but is grown to a small extent in the southwestern United States.

Seaweed gums produced by hot water or alkaline extraction of seaweed are other important industrial raw materials.

Industrially important or potentially useful gums are listed in the accompanying table, together with the plant source, the sugar units, and the most prevalent glycosidic linkage when known.

Sources and sugars of important gums

Gum	Source	Sugars present and linkages
Seaweeds		
Agar	Red algae (*Gelidium* sp.)	D-galactose β-(1 → 4), 3,6-anhydro-L-galactose α-(1 → 3), + sulfate acid ester groups
Algin	Brown algae (*Macrocystis pyrifera*)	D-mannuronic acid β-(1 → 4), L-guluronic acid β-(1 → 4), Na salt
Carrageenin	Red algae (*Chondrus crispus*) (*Gigartina stellata*)	D-galactose, 3,6-anhydro-D-galactose + sulfate acid ester groups
Fucoidan	Brown algae (*Fucus* sp., *Laminaria* sp.)	L-fucose + sulfate acid ester groups
Laminaran	Brown algae (*Laminaria* sp.)	D-glucose, D mannitol, β-(1 → 3) chain and β-(1 → 6) branches
Plant exudates		
Gum arabic	*Acacia* sp	L-arabinose, D-galactose, L-rhamnose, D-glucuronic acid
Ghatti	*Anogeissus latifolia*	L-arabinose, D-xylose, D-galactose, D-mannose, D-glucuronic acid
Karaya	*Sterculia urens*	D-galactose, L-rhamnose, D-galacturonic acid
Tragacanth	*Astragalus* sp.	D-galactose, D-xylose, D-glucuronic acid
Plant extracts		
Pectin	Cell walls and intracellular spaces of all plants (commercial source, citrus waste)	D-galacturonic acid α-(1 → 4) partially esterified, L-arabinose α-(1 → 5) and α-(1 → 3) branches, D-galactose β-(1 → 4)
Larch arabinogalactan	Western larch	D-galactose, L-arabinose
Ti	Tubers of *Cordyline terminalis*	D-fructose, D-glucose
Plant seeds		
Corn-hull gum	Corn seed-coat	D-xylose, L-arabinose, D-galactose, L-galactose, D-glucuronic acid
Guar	*Camposia teragonolobus* endosperm	D-mannose β-(1 → 4), D-galactose α-(1 → 6) branches
Locust bean	Carob tree (*Ceratonia siliqua*) endosperm	D-mannose β (1 → 4), D-galactose α-(1 → 6) branches
Quince seed	*Cydonia vulgaris*	L-arabinose, D-xylose, hexuronic acid, monomethyl hexuronic acid
Psyllium seed	*Plantago* sp	D-xylose, L-arabinose, D-galacturonic acid, L-rhamnose, D-galactose

Sources and sugars of Important gums (Cont.)

Gum	Source	Sugars present and linkages
Flax seed	*Linum usitatissimum*	D-galacturonic acid, D-xylose L-rhamnose, L-arabinose, L-galactose, D-glucose
Tamarind	*Tamarindus indica*	D-glucose, D-galactose, D-xylose
Wheat gum	Wheat	D-xylose β-(1 $\rightarrow$ 4), L-arabinose branches
Miscellaneous		
Cellulose derivatives	Plant cell walls, wood pulp and cotton	D-glucose β-(1 $\rightarrow$ 4)
Starch	Cereal grains and tubers	
Amylose		D-glucose α-(1 $\rightarrow$ 4)
Amylopectin		D-glucose α-(1 $\rightarrow$ 4), α-(1 $\rightarrow$ 6) at branch points
Dextran	Bacterial action on sucrose	D-glucose α-(1 $\rightarrow$ 6) and α-(1 $\rightarrow$ 3)
Chitin	Exoskeleton of animals of the phylum Arthropoda	*N*-acetyl-D-glucosamine β-(1 $\rightarrow$ 4)

Use. Gums are used in foods as stabilizers and thickeners They form viscous solutions which prevent aggregation of the small particles of the dispersed phase. In this way they aid in keeping solids dispersed in chocolate milk, air in whipping cream, and fats in salad dressings. Gum solutions also retard crystal growth in ice cream (ice crystals) and in confections (sugar crystals). Their thickening and stabilizing properties make them useful in water-base paints, printing inks, and drilling muds. Because of these properties they also are used in cosmetics and pharmaceuticals as emulsifiers or bases for ointments, greaseless creams, toothpastes, lotions, demulcents, and emollients. Gums are used to modify texture and increase moisture-holding capacity, for example, as gelling agents in canned meats or fish, marshmallows, jellied candies, and fruit jellies. Their adhesive properties make them useful in the production of cardboard, postage stamps, gummed envelopes, and as pill binders. Other applications include the production of dental impression molds, fibers (alginate rayon), soluble surgery films and gauze, blood anticoagulants, plasma extenders, beverage-clarifying agents, bacteriological culture media, half-cell bridges, and tungsten-wire-drawing lubricants. *See* FOOD MANUFACTURING. [J.R.Y.; R.L.WH.]

Bibliography: American Chemical Society, *Natural Plant Hydrocolloids*, Advances in Chem. Ser., no. 11, 1954; C. L. Mantell, *The Water Soluble Gums*, 1947; W. Pigman and M. L. Wolfrom (eds.), *Advances in Carbohydrate Chemistry*, vol 13, 1958; R. L. Whistler and J. BeMiller (eds.), *Industrial Gums: Polysaccharides and Their Derivatives*, vol. 1, 1959; R. L. Whistler and C. L. Smart, *Polysaccharide Chemistry*, 1953.

Gun sights

Optical instruments which establish an optical line or axis for the purpose of aiming a weapon. The

reflective mirror.

Rifle sights. A typical rifle sight (Fig. 1) comprises a terrestrial telescopic system having an objective eyepiece, erector lens, and reticle (*see* TELESCOPE). Sometimes a field lens is employed to ensure uniform illumination. The magnifying powers customarily range from unity to 12 diameters, magnifications of 2½–9 being most common. Instruments are available in which the power may be varied optically over a wide range to suit existing conditions. Large eyepieces are used to provide a wide field of view and a substantial eye relief, the latter so that the eyepiece will not strike the user's eye in recoil. In order to adjust for elevation or windage, either the reticle may be moved, usually by means of screws, or the entire sight may be tilted relative to the gun barrel. Machine gun sights are also telescopic in nature and customarily employ a prism for erecting the image. Magnifying powers range between 1½ and 3, and the reticle may be illuminated by a small lamp to permit night use.

Aircraft gun sights. These are usually of the reflector type (also known as reflex sights) and employ in their simplest form a lamp, reticle, collimating lens, and a glass plate or partially reflecting mirror (Fig. 2). The collimator images the reticle pattern at infinity and the mirror superimposes this image over the target area. The collimator's

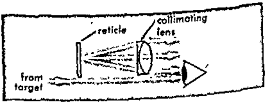


Fig. 1. Rifle sight.

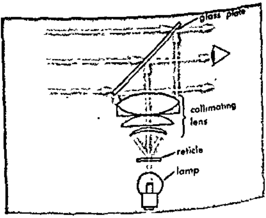


Fig. 2. Reflex sight.

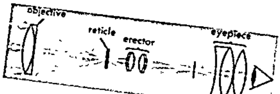


Fig. 3. Collimator sight.

pupil is so large that the eye can be positioned at any point within a large cylindrical volume in back of the sight. The collimating system may employ a simple spherical mirror, a Schmidt lens, or a Mangin mirror. A Mangin mirror consists of a negative meniscus lens, one surface of which is silvered to act as a spherical mirror, while the other surface corrects for the spherical aberration of the reflecting surface (*see* MIRROR OPTICS). The magnification of this type of sight is unity.

Artillery sights. These can assume various forms, the simplest of which is the collimator sight (Fig. 3), consisting of an objective having a reticle at its focus. When the eye is so placed as to receive light simultaneously from the target and the reticle, the latter appears superimposed on the former and a line of sight is established. Other forms of artillery sights include the elbow telescope, which is a conventional terrestrial telescope containing a prism for producing a right angle bend, and the panoramic sight. For a discussion of panoramic sights, *see* PERISCOPE; *see also* LENS, OPTICAL.

[E.K.K.]

Bibliography: C. H. von Hofe, *Fernoptik*, 1941; D. H. Jacobs, *Fundamentals of Optical Engineering*, 1943.

Gunite

A trade name denoting a method for placing cement mortar by pneumatic means. Gunite is also called shotcrete. A dry mixture of cement and sand about 1:4 ratio by volume is blown by air jet through a nozzle so as to impinge against a construction form or other surface. Enough water is added at the nozzle to give cohesion at the point of impact. Compaction is provided by the impact. Gunite is used in repairing structures, for protective coverings of many types (such as fireproofing), and for relatively thin reinforced walls. Its advantages are low form cost (often no forms are needed) and its ability to produce difficult shapes easily. Its disadvantages are the need for skilled workmen and the loss of material and other complications caused by the rebounding of material from the surface. Shrinkage is relatively high. *See* PORTLAND CEMENT.

[I.L.T.]

Guppy

The most common of the tropical aquarium fishes, *Lebistes reticulatus*; also the mosquito fish, *Gambusia affinis*. *Lebistes* is a native of Trinidad and Venezuela, whereas *Gambusia* ranges from New Jersey to Mexico. *Lebistes* is kept in aquariums throughout the world. The guppy gives birth to living young, producing as many as 20–25 young every 28 days, and is a very easy fish to breed and rear in captivity. The females reach a length of 2 in. and are brown in color, in sharp contrast to the 1-in. males which are marked in a variety of black,

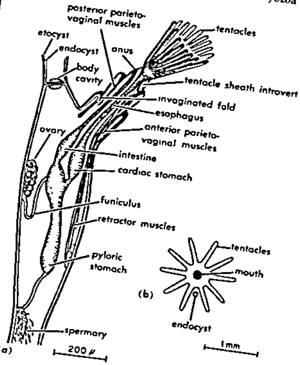
The guppy, Gambusia affinis; length to 2¼ in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

red, orange, yellow, and blue spots and lines. Some highly developed strains are almost entirely red; others are golden; and still others have special markings or patterns on the tails. *See* CYPRINODONTIFORMES.

[J.D.B.]

Gymnolaemata

A class of the Ectoprocta which contains about 155 families, 1256 genera and subgenera, and thereby most of the species of the phylum Bryozoa



(a) Paludicella articulata, a ctenostomatous bryozoan. Detail of zooid whose tentacles are evaginated. (b) Bryozoan tentacular crown of Bowerbankia, a gymnolaematous ectoproct.

(R. Bassler 1953). Most are marine species except for a few which are brackish and fresh water. The class occurs from the Ordovician Period to the present. Its species have a circular lophophore with approximately 8–35 tentacles encircling a mouth devoid of an overhanging fold, the epistome. *See* BRYOZOA. [M.D.RO.]

Gymnophiona

The smallest order of the class Amphibia known as the caecilians. These are wormlike, legless animals with indistinct or even hidden eyes. A series of annular grooves is usually present along the length of the body, heightening the resemblance to an earthworm. Most caecilians lead a burrowing existence, though members of one genus, *Typhlonectes*, are aquatic. Some species have the eyes hidden beneath the bones of the skull and probably are blind, but others at least are able to distinguish movement. A unique feature of the caecilians among modern Amphibia is the presence of scales buried in the skin of some species.

There are approximately 70 species of caecilians confined to tropical regions of both eastern and western hemispheres. Many species are less than 1 ft in length, but three species of the genus *Caecilia* grow to over 3 ft. The breeding habits of caecilians are poorly known. Some species lay eggs, while others bring forth their young alive. The embryos of the species that bear living young are nourished in the later part of their embryonic development by "uterine milk" secreted by the mother In at least some of the species that lay eggs there is an aquatic larval stage. Caecilians are carnivorous, but little is known of the food habits except that captive specimens have fed on earthworms. *See* AMPHIBIA; CAUDATA; SALIENTIA. [R.C.Z.]

Gymnospermae

A class of the plant subphylum Pteropsida, and the most primitive of the seed-bearing plants. During the early Mesozoic, the gymnosperms were the most advanced plants in the evolutionary scale and formed the dominant vegetation. Today there are less than 1000 species, but despite this decline, the gymnosperms still continue to be an important group of plants All are woody, mostly trees, and nearly all of them are evergreen. Unlike angiosperms, the gymnosperms, with few exceptions, have no true vessels in the vascular (water-conducting) system Some species form extensive forests of an almost pure stand which are sources of valuable timber.

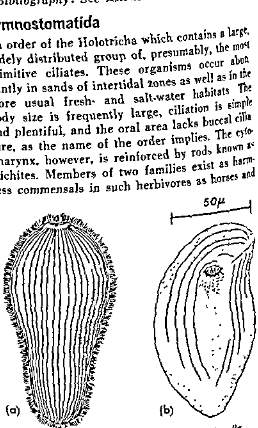
Reproduction. In reproduction, the large independent foodmaking sporophyte (sporeproducing generation) alternates with the small dependent gametophyte (sperm- and egg-producing generation). There are two kinds of spores (heterospory): microspores (pollen) which produce the male gametophytes, and megaspores from which the female gametophytes develop within the megasporangium (ovule). Usually, the female gametophyte produces several archegonia (egg-containing structures) Thus, following fertilization several embryos appear in a single ovule, but only one survives. The megasporangia and, subsequently, the seeds which are mature ovules are exposed on the surface of the megasporophyll (scales), hence the name gymnosperms meaning "naked seeds." There is no highly specialized flower such as appears in angiosperms and in most gymnosperms the microspores and megaspores are borne in strobili (cones).

Classification. The gymnosperms are classified in seven orders: Cycadofilicales, Bennettitales, Cycadales, Cordaitales, Ginkgoales, Coniferales (the pines and related species), and Gnetales Of these orders the Cycadofilicales, Bennettitales, and Cordaitales are extinct and are known only through the study of their fossil remains. See separate articles describing each of these orders; *see also* PALEOBOTANY; PLANT KINGDOM. [P.D.S.]

Bibliography: See EMBRYOPHYTA.

Gymnostomatida

An order of the Holotricha which contains a large, widely distributed group of, presumably, the most primitive ciliates. These organisms occur abundantly in sands of intertidal zones as well as in the more usual fresh- and salt-water habitats The body size is frequently large, ciliation is simple and plentiful, and the oral area lacks buccal ciliature, as the name of the order implies. The cytopharynx, however, is reinforced by rods known as trichites. Members of two families exist as harmless commensals in such herbivores as horses and camels and in a few other mammals. *Prorodon*, (illustration *a*), *Holophrya*, *Didinium*, and *Dileptus* are examples of carnivorous gymnostomes. *Chilodonella* (illustration *b*) and *Nassula* are herbivorous and anatomically more complex. *Ichthyophthirius*, a fish parasite, is often classified erroneously as a gymnostome. In reality, it belongs in the order Hymenostomatida. It is found as a parasite on fish in home aquaria and causes the disease known as the "ich." *See* HOLOTRICHA. [J.O.C.]



Gymnostomatida. (*a*) Prorodon. (*b*) Chilodonella

Gynogenesis

The development of a fertilized egg through the action of the egg nucleus, without participation of the sperm nucleus. However, the spermatozoon penetrates the egg, as it does in normal fertilization, and usually contributes the division center (centrosome) for the establishment of a bipolar division apparatus.

Gynogenesis occurs as a regular phenomenon in nature in some nematodes, flatworms, other invertebrates, and in some species of fishes. In these cases, maturation of the egg is usually modified to prevent reduction of the chromosome number, and gynogenetic development is diploid.

Gynogenesis also occurs spontaneously in species that usually have normal fertilization, and produces haploid embryos. It can be induced experimentally by heavy irradiation or chemical treatment of the sperm fluid. These procedures selectively damage the nuclear material of the spermatozoa. *See* ANDROGENESIS; CYTOLOGY; FERTILIZATION; MEROGONY. [C.FA.]

Gypsum

A mineral with the chemical composition $CaSO_4 \cdot 2H_2O$. Gypsum is the commonest of the sulfate minerals. It occurs in five varieties: (1) rock gypsum; (2) gypsite, an impure earthy form; (3) alabaster, a massive, fine-grained, translucent variety; (4) satin spar, a fibrous silky form; and (5) selenite, transparent crystals. Crystals of gypsum are monoclinic, clear, white to gray, yellowish and brownish in color, with well-developed cleavages. Platy fragments are flexible but not elastic. Luster is subvitreous to pearly. Hardness is 2 on Mohs scale and specific gravity is 2.3. It fuses readily. It is soluble in hydrochloric acid and slightly soluble in water. *See* CALCIUM.

Gypsum is calcined in kettles or kilns at temperatures of 190–200°C to remove part of the water of crystallization. The calcined product, known as plaster of paris, sets in 6–8 min after addition of water. The setting time can be extended to 1–2 hours by adding

portland cement. Ground gypsum and anhydrite are used in the southern United States as soil conditioners to improve permeability Gypsum and anhydrite, particularly the latter, may be made into ammonium sulfate fertilizer. Calcined gypsum is also used for industrial plasters, such as those used in pottery, molding, dentistry, and statuary. Pure white uncalcined gypsum, known as terra alba, is used as a filler in paper and paints and as a nutrient in growing yeast. Other uncalcined ground gypsum is used as a filler in insecticide. *See* ANHYDRITE.

Gypsum is deposited from sea water or brines from salt lakes at temperatures below 42°C as bedded deposits. It is also precipitated from aqueous solutions in limestone caves, such as the famous gypsum flowers of Mammoth Cave, Kentucky. Large deposits of gypsum are generally associated with limestone, dolomite, shale, sandstone, and salt beds. *See* EVAPORITE (SALINE).

Gypsum is of world-wide distribution. Thick commercial deposits of calcium sulfate occur in Nova Scotia and in Stassfurt, Germany. In the western United States extensive commercial gypsum deposits occur in California, Nevada, Utah, Texas, Iowa, Oklahoma, Kansas, Montana, Colorado, Arizona, South Dakota, Wyoming, and New Mexico. In the East, large mines are located in Michigan, New York, Ohio, and Indiana. [E.C.T.C.]

Gypsum plank

A structural, precast roof deck unit. All four edges are bound with a steel framework of tongue-and-groove design allowing adjacent units to lock to form a series of rigid I beams. The gypsum core is reinforced with welded galvanized-steel mesh. The planks are clipped to supporting steel purlins and topped with a built-up roof covering. The advantages of gypsum plank are fire-resistance, light weight, and the ease with which it is erected. [C.N.G.]

Gyrator

A linear, passive, four-pole, electric circuit element whose transmission properties are such that it is effectively a half wavelength longer for one direction of transmission than for the other direction of transmission. In other words, a gyrator is a device that causes a reversal of signal polarity for one direction of propagation but not for the other direction of propagation. (A four-pole element has a pair of input terminals, or poles, and a pair of output terminals.) The device is novel, since it violates the theorem of reciprocity. *See* RECIPROCITY, PRINCIPLE OF.

Until a few years ago, all known linear passive electric networks obeyed the theorem of reciprocity (*see* NETWORK THEORY, ELECTRICAL). Today several different types of passive nonreciprocal microwave networks are in practical use. Through the use of these devices it is now possible to build so-called one-way transmission systems. These microwave systems can propagate a wave in one direction with negligible attenuation but almost completely absorb a wave traveling in the opposite direction through the system. In addition, it is now possible to perform the operation of duplexing without power loss. In other words, the ability to violate reciprocity allows one, in principle at least, to send and receive simultaneously the same frequency from the same antenna. *See* CONTINUOUS-WAVE RADAR.

Reciprocity. The theorem of reciprocity can be stated in many equivalent forms, and perhaps the simplest to consider and understand is the particular form which it takes when it is expressed especially for a two-terminal microwave network. Thus, if the box in Fig. 1 represents such a microwave

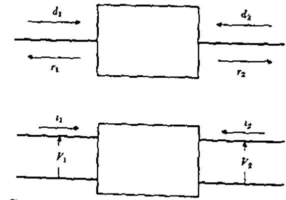


Fig. 1. Scattering matrix representation of a two-terminal microwave network.

network, where d_1 represents the amplitude and phase of the wave incident on the network from the left, r_1 represents the wave propagating out from the network on the same side, and d_2 and r_2 represent the same quantities on the other side, it is relatively easy to see what the theorem of reciprocity means.

In a linear passive network both r_1 and r_2 are made up of two components. Thus r_1 consists of a part which arises from a partial reflection of d_1 at the input terminals of the network and a part consisting of that fraction of d_2 which is transmitted through the network.

$$\begin{aligned} r_1 &= S_{11}d_1 + S_{12}d_2 \\ r_2 &= S_{21}d_1 + S_{22}d_2 \end{aligned} \tag{1}$$

The significance of the above coefficients becomes obvious if one lets d_1 and d_2 be separately equal to zero. Thus when d_2 is zero

$$S_{21} = \left(\frac{r_2}{d_1}\right)_{d_2=0} \tag{2}$$

In other words, S_{21} gives the amplitude and phase of the wave emerging on the right-hand side of the network when a unit wave, with zero phase angle, is incident on the left-hand side. Thus if the network has a zero insertion loss for propagation from left to right, and if the input and output terminals are chosen to be electrically at such a plane that the phase of r_2 is identical to that of d_1, then the quantity S_{21} would be equal to unity. The significance of all the other coefficients above should be obvious to the reader by similar arguments. If the network is perfectly matched when looking in both directions, then there are no reflections, and S_{11} and S_{22} are zero. These coefficients turn out to be very useful in the analysis and design of microwave networks and are referred to as the scattering matrix of the network.

The theorem of reciprocity demands that

$$S_{12} = S_{21} \tag{3}$$

In other words, the network has the same transfer characteristic for one direction of propagation as it has for the other. Thus, if a matched network has a certain insertion loss or causes a certain phase shift for one direction of propagation, then it must have the same insertion loss and phase shift for the other direction of propagation. An ideal gyrator is defined as the microwave device in which S_{11} and S_{22} are zero and S_{12} is equal to the negative of S_{21}.

For the reader who is more familiar with ordinary four-pole circuit analysis than he is with microwave networks, the reciprocity theorem can be equally well stated. Equation (1) would be written

$$\begin{aligned} V_1 &= Z_{11}i_1 + Z_{12}i_2 \\ V_2 &= Z_{21}i_1 + Z_{22}i_2 \end{aligned} \tag{4}$$

The theorem of reciprocity can be simply stated for a four-pole network by stating the condition

$$Z_{12} = Z_{21} \tag{5}$$

Theoretical gyrators. The first comprehensive treatment of antireciprocal four-pole networks was given in 1948 by B. D. H. Tellegen, who coined the word gyrator to describe such networks. Tellegen restricted his analysis to a so-called ideal gyrator

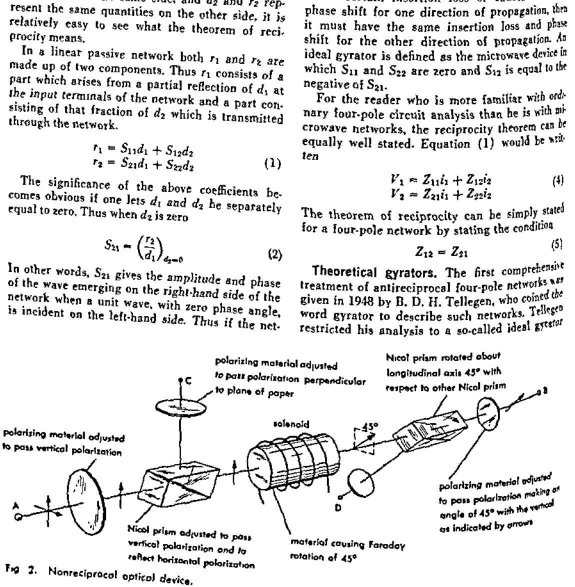


Fig. 2. Nonreciprocal optical device.

in which the impedance matrix had the following form:

$$Z_{11} = Z_{22} = 0 \tag{6}$$

$$Z_{12} = -Z_{21} = -R \tag{7}$$

Now it is easily shown that, for such a gyrator to be nondissipative, the transfer impedances must be real, as Tellegen has shown them; however, it is not necessary that the diagonal terms Z_{11} and Z_{22} of the impedance matrix be zero. In addition, the restriction imposed by Eq. (6) is also not necessary in order to realize two of the most important properties of the gyrator, which are the construction of one-way transmission systems and the phase inversion of signal polarity for one direction of transmission. As a result of the restriction placed on the so-called ideal gyrator by Eq. (6), Tellegen's gyrator has the additional interesting property of impedance inversion. Thus if Tellegen's gyrator is terminated by an impedance

$$Z_L = \frac{V_2}{i_2}$$

the input impedance of the gyrator from Eq. (4) is

$$Z_{\text{in}} = \frac{V_1}{i_1} = R^2/Z_r \tag{8}$$

This property of impedance inversion is not, however, a distinguishing property of an antireciprocal network; for if one makes the same requirement in a nondissipative reciprocal four-pole, that $Z_{11} = Z_{22} = 0$, the input impedance of the terminated network is found to be

$$Z_{\text{in}} = \frac{V_1}{i_1} = \frac{X_{12}^2}{Z_L} = \frac{Z_{12}Z_{21}}{Z_L} = \frac{Z_{12}^2}{Z_L} \tag{9}$$

Such a reciprocal impedance-inverting network can be easily realized.

Practical gyrators. The theorem of reciprocity has, in the past, been considered so universally valid that present-day textbooks still make the statement that, if the condition stated in Eq. (5) is valid, the four-pole is passive, and that, if the condition does not hold for a particular network, the network must of necessity be an active network.

Actually the theorem of reciprocity is as universally valid in mechanical, acoustical, or optical systems as it is in electrical systems; but it is known that passive systems can be built in all these fields which violate the theorem of reciprocity. These systems are so unique, however, that they have obtained special attention in the technical literature for over fifty years. Thus, it has been known for some time that a mechanical system which contains a gyroscopic coupler does not obey the theorem of reciprocity, and as early as the turn of the century, scientists were worried about the apparent lack of reciprocity which arose in the operation of certain electromechanical transducers.

Probably one of the first passive nonreciprocal systems proposed in the literature was an optical one-way transmission system, proposed by Lord Rayleigh, making use of the Faraday rotation of the plane of polarization of light. This effect, first discovered by Faraday, is exhibited by all transparent substances. Thus, if a polarized light ray is propagated through a transparent substance and in the direction of a superimposed magnetic field, the plane of polarization of the light is rotated through some angle θ which is determined by the characteristics of the substance and the strength of the magnetic field. The Faraday rotation is unusual in that it is nonreciprocal. Thus the rotation is in the same direction regardless of the direction of propagation. Hence if the plane of polarization of a light ray is rotated through an angle θ in traversing the Faraday cell and the ray is then reflected back to the source, the polarization is rotated again through an angle θ so that when it arrives back at the source the ray will have its plane of polarization rotated through a total angle of 2θ.

Lord Rayleigh's one-way system shown in Fig. 2 consisted of two polarizing Nicol prisms oriented so that their planes of acceptance made an angle of 45° with each other and with the material causing the Faraday rotation placed between them. Thus, light which was passed by the first crystal and whose plane of polarization was rotated 45° would be passed by the second crystal also. In the reverse direction, however, the rotation is in such a sense that light admitted to the system by the second crystal would not be passed by the first. This light would be reflected by the Canada balsam cement in the Nicol prism on the left side and would be directed toward point C in the figure. Thus, light admitted to the device at point A would be transmitted to point B; light admitted at point B would be transmitted to point C; light admitted at point C would be transmitted to point D; and light admitted at point D would be transmitted to point A.

The microwave analog of Lord Rayleigh's device was first proposed by C. L. Hogan and is shown in Fig. 3. Since it circulates power from wave guide A to B, from B to C, from C to D, and from D to A, it has been called a circulator. An interesting property of the circulator could be demonstrated if one could attach a perfectly matched antenna to arm B. Then if a transmitter were attached to arm A and a receiver to arm C, it would be possible to send and receive simultaneously the same frequency from the same antenna. This feat has actually been achieved, but it is difficult to build an antenna that is matched over a wide enough bandwidth to make such an application of great practical interest.

If, however, a transmission line is attached at point B, a practical duplexer, which is capable of separating signals traveling in opposite directions on the transmission line, can be achieved. If arms C and D of Fig. 3 are terminated with matched attenuators and arms A and B are connected to a mi-

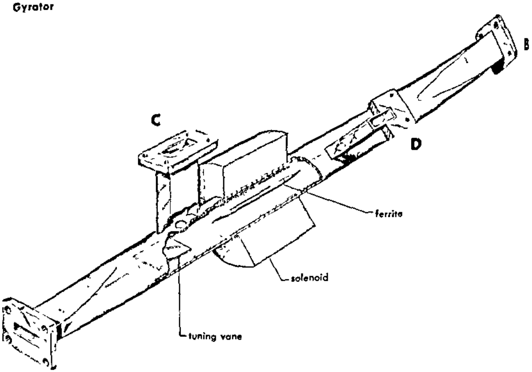


Fig. 3. Microwave analog of nonreciprocal optical device.

crowave system, then a one-way transmission system will result. That is, power entering from A would propagate through arm B to the microwave system, but power entering from B would not propagate back through the system connected at A but instead would be absorbed by the attenuator in arm C. A variation of this system is shown in Fig. 4.

If the magnetic field imposed on the rotating section is caused to vary with time, then the angle of rotation will also vary with time, and the amount of power propagating from A to B will be modulated by the magnetic field. Simple microwave modulators can thus be built from this basic element, and although they do not depend on the nonreciprocal property of the device for their operation, they are nevertheless worth mentioning here.

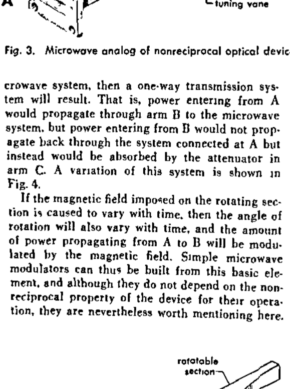


Fig. 4 Practical nonreciprocal microwave device.

The only practical gyrators, circulators, and isolators in use today are derived from the devices discussed above and use ferrites in wave-guide systems. However, many proposals have been made for achieving these devices by other approaches. These other approaches are of interest because they might give the possibility of building nonreciprocal devices at lower frequencies than has yet been possible using ferrites.

In 1946, E. M. McMillan showed that an antireciprocal four-pole electrical network could be obtained by properly coupling electromechanical transducers. Although this proposal has not yet led to a practical device, it is of interest because it was probably the first proposal for realizing a passive electrical system which violated the reciprocity theorem.

A gyrator proposed by L. Goldstein and M. A. Lampert arises from the behavior of electrical charges moving in a magnetic field. They proposed replacing the ferrite used by Hogan with a bottle of gas, which is kept constantly ionized by means of a glow discharge. This particular device generates a great deal of microwave noise and has not come into practical use. A natural extension of their work, however, is to use the free charges (holes and electrons) which exist in a semiconductor rather than in an ionized gas. In this case, the rotation of the plane of polarization of the electric field is related to the Hall effect in a semiconductor. Such a device must be lossy and hence will probably never compete with ferrite gyrators at frequencies at which ferrite gyrators can be built. However, Hall effect gyrators offer interesting possibilities for low frequency operation in spite of their insertion loss.

An interesting potential application of such devices is to place them between negative-resistance devices to allow the cascading of such devices so that the gain can be increased without oscillation. To date, one of the biggest drawbacks preventing the use of negative-resistance devices, such as Esaki (tunnel) diodes, for amplification has been the inability to cascade these devices in an amplifier.

Much has been written about the conditions under which it might be possible to build an antireciprocal element such as the gyrator. It has been shown that in order for a completely electrical passive system to violate reciprocity, a nonchanging magnetic field must be present. Only two types of practical gyrators have been built to date. Needless to say, all practical proposals made for realizing a gyrator satisfy the requirement of having a dc-biasing magnetic field acting on them. *See* FERRITE DEVICES. [C.L.H.]

Bibliography: A. Bloch, A new approach to the dynamics of systems with gyroscopic coupling terms, *Phil. Mag.*, [7]35(244):315–334, 1944; H. B. G. Casimir, On Onsager's principle of microscopic reversibility, *Revs. Mod Phys.*, 17(2–3):343–350, 1945; L. Goldstein and M. A. Lampert, New linear passive non-reciprocal microwave circuit component, *Proc. IRE*, 41(2):295–296, 1953; C. L. Hogan, The microwave gyrator, *Bell System Tech. J.*, 31(1):1–31, 1952; E. M. McMillan, Violation of the reciprocity theorem in linear passive electromechanical systems, *J Acoust. Soc. Amer.*, 18(2):344–347, 1946; Lord Rayleigh, On the magnetic rotation of light and the second law of thermodynamics, *Nature*, 64(1667):577–578, 1901; B. D. H. Tellegen, The gyrator, a new electric network element, *Philips Research Repts.*, 3(2):81–101, 1948.

Gyrocompass

A north-seeking form of gyroscope used as a shipboard directional reference. The first practical gyrocompasses were developed by H. Anschütz

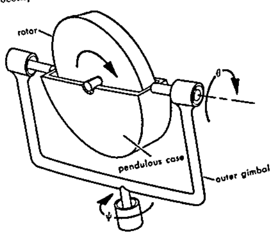


Fig. 1. Gyrocompass model.

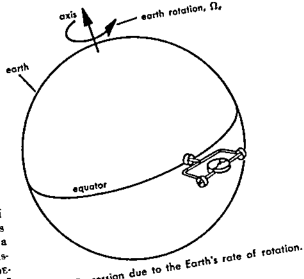


Fig. 2. Precession due to the Earth's rate of rotation.

(Germany) in 1908, E. A. Sperry (United States) in 1911, and S. G. Brown (England) in 1916. Modern gyrocompasses are so reliable and so much more accurate than magnetic compasses that they are now used as the prime navigational instrument on nearly every ship.

Basic operation. A gyrocompass combines the action of two devices, a pendulum and a gyroscope, to produce alignment with the Earth's spin axis. The principle is demonstrated with the model of Fig. 1, which consists of (1) a rapidly spinning, heavy gyro rotor, (2) a pendulous case which permits the rotor axle to nod up and down (angle θ), and (3) an outer gimbal which permits the axle to rotate in azimuth (angle ψ).

In Fig. 2 the model is positioned at the equator of the Earth. As the Earth rotates, the gimbal moves with it. So long as the rotor's spin axis is aligned with the Earth's axis, the gyro experiences no torque from Earth rotation. If there is misalignment, however, a sequence of restoring torques is initiated.

The restoring action can be seen by imagining a beam of light to be projected by the gyro rotor axle onto a vertical piece of paper held just at the north end of the axle. Figure 3*a* shows a typical pattern traced on the paper by the light spot. At point A the rotor axle has been displaced east by $\psi = 5°$. The spin axis of the gyro is no longer parallel to that of the Earth. Therefore, as the Earth rotates, the north end of the gyro rotor tilts upward. The ensuing dip angle results in a pendulum torque, down at the south end and up at the north end of the rotor axis. This torque causes the north end of the gyro axis to precess westward. As soon as the gyro axis passes through the north-south meridian, the action reverses. The north end of the rotor tilts down and the pendulum torque causes the gyro to precess eastward. Thus, the axis traces the elliptic path in Fig. 3*a*. In practical instruments damping is added and the path con-

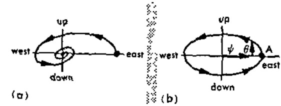


Fig 3. Path of rotor axle of gyrocompass. (a) Typical undamped pattern. (b) With damping added.

verges to true north in a spiral motion as in Fig. 3b. See GYROSCOPE

In a shipboard installation the system must be mounted in a complete set of gimbals to isolate it from rolling, pitching, and yawing motions of the ship. Friction must be minimized. Moreover, Schuler tuning is employed to keep horizontal accelerations of the ship from producing false torques on the pendulum; the unique combination of gyro spin speed and pendulosity is chosen so that no acceleration of the instrument can disturb its vertical reference. Then the light spot in Fig. 3 encircles the true north point once every 84 min. See SCHULER PENDULUM.

If the ship travels north or south the ship's velocity V over the Earth creates an apparent Earth rate of $\Omega_v = V/R$ (R is the Earth's radius) about an axis perpendicular to the Earth's spin axis. The gyrocompass aligns to the vector sum of Ω_e and Ω_v, which results in a north steaming error angle whose tangent is Ω_v/Ω_e. This error is corrected by using V from the ship's pitometer to precess the gyro in the opposite direction at an equal rate.

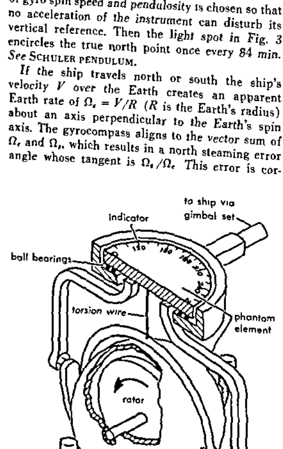


Fig 4. Typical gyrocompass design.

Practical gyrocompass design is shown by the typical, widely used system depicted in Fig 4. The rotor and case of Fig. 4 replace the rotor and case of Fig. 1, except that in Fig. 4 this assembly is not itself pendulous. Instead the case carries a coupling pin which rides in a slot on the ballistic ring, which is pendulous by virtue of the two mercury bottles and connecting tube known as a mercury ballistic. Gyroscopic damping is introduced by making the coupling pin slightly eccentric.

The ballistic ring is supported by bearings from the phantom element, which replaces the outer gimbal of Fig. 1. The gyro assembly is also supported from the phantom element by a torsion wire which provides frictionless support. A servo motor is arranged to drive the phantom element whenever there is any twist in the wire. Thus the phantom element always stays aligned with the gyro. The phantom element carries an indicator from which ship's heading is read. [RHC]

Bibliography: K. I. T. Richardson, *The Gyroscope Applied*, 1954.

Gyrocotylidea

An order of tapeworms of the subclass Cestodaria. All species are intestinal parasites of chimaeroid fishes and have an anterior end with an eversible proboscis and a posterior end with a ruffled adhesive organ. Typically, only one sexually mature worm is found in an individual host The life history is incompletely known, although it has been claimed that, unlike other tapeworms, gyrocotylids can develop in a single host without larval development in some other host species. See CESTODARIA [CPR]

Gyromagnetic effect

An effect arising from the relation between the angular momentum and the magnetization of a magnetic substance. It is the effect which is exploited in the measurement of the gyromagnetic ratio of magnetic materials (see GYROMAGNETIC RATIO). The gyromagnetic effect is demonstrated by a simple experiment in which a freely suspended magnetic substance is subjected to a magnetic field. Upon a change in direction of the magnetic field, the magnetization of the substance must change. In order for this to happen, the atoms must change their angular momentum. Since there are no external torques acting on the system, the total angular momentum must remain constant. Thus, the sample must acquire a mass rotation which may be measured. In this way, the gyromagnetic ratio may be determined. Two common methods of determination are the Einstein-de Haas method and the Barnett method.

Einstein–de Haas method. This is usually used to determine the gyromagnetic ratio of ferromagnetic materials. Imagine a ferromagnetic substance

n the shape of a cylinder suspended on one end by a torsion fiber, forming thereby a torsional pendulum. A magnetic field is applied parallel to the axis of the cylinder, for example, by means of a coil surrounding the cylinder. The ferromagnetic cylinder is now magnetized in the direction of the magnetic field. If the field is suddenly reversed, the magnetization will reverse, and the accompanying change in angular momentum of the atoms will be balanced by a mass rotation of the cylinder which can be measured by noting the change in amplitude of displacement of the torsional pendulum. A measurement is made of the change in magnetization by means of a magnetometer. The ratio of magnetization change to angular momentum change is the magnetomechanical ratio, the reciprocal of the gyromagnetic ratio.

Barnett method. This is an alternative to the Einstein–de Haas technique. In this experiment, a ferromagnetic bar with a coil surrounding it is spun rapidly about its axis. The atoms, therefore, acquire an angular momentum in the direction of rotation. The magnetization thereby produced is measured by stopping the rotation abruptly and measuring the voltage induced in the surrounding coil which is part of a fluxmeter circuit. Now the same increase in magnetization in the direction of the rotation could be achieved by applying a magnetic field in that direction instead of rotating the bar. In the first case, if an angular momentum L makes an angle θ with the axis of rotation, it will experience a torque $L\omega \sin\theta$, where ω is the angular velocity of rotation, tending to turn it into the axis of rotation. In the second case, the magnetic field H would produce a torque on L in the same direction of $(1/\gamma)LH \sin\theta$, where $1/\gamma$ is the magnetomechanical ratio (L/γ = magnetic moment). If the magnetizations in the two cases are the same, then the torques are the same also, and

$$L\omega \sin\theta = (1/\gamma)LH \sin\theta$$

Thus

$$\gamma = H/\omega$$

In this way, the magnetomechanical ratio can be measured. In an experiment similar to the one described, S. J. Barnett showed in 1914 that the magnetomechanical ratio for a ferromagnet was close to e/mc, the ratio of charge to mass of an electron. [E.A.; F.K.]

Bibliography: S. J. Barnett, *Proc. Am. Acad. Arts Sci.*, 75:109, 1944; C. Kittel, *Introduction to Solid State Physics*, 2d ed., 1956

Gyromagnetic ratio

The ratio of angular momentum to magnetic moment for atomic systems. This ratio is usually expressed in terms of the magnetomechanical factor g':

$$\text{angular momentum/magnetic moment} = 2mc/g'e$$

The ratio is written here in electromagnetic units; thus, e/c and m are the charge and mass of the electron. The factor g' is sometimes loosely called the gyromagnetic ratio.

Magnetomechanical ratio. This quantity is the inverse of the gyromagnetic ratio. It is usually denoted by γ and is equal to $g'e/2mc$.

The magnetomechanical ratio of a substance identifies the origin of the magnetic moment. For example, for electron spin, the angular momentum is $\frac{1}{2}\hbar$, where $\hbar$ is Planck's constant divided by 2π (see ELECTRON SPIN). The magnetic moment is the Bohr magneton $e\hbar/2mc$. Thus, the magnetomechanical ratio is

$$\gamma = \frac{e\hbar/2mc}{\hbar/2} = e/mc$$

Since $\gamma = g'e/2mc$, for electron spin $g' = 2$. For orbital angular momentum, $\gamma = e/mc$ and $g' = 1$. The experimental values of g' for most ferromagnetic materials are in the neighborhood of 2, showing that the major contribution to the magnetization comes from the electron spin. Deviations of g' from 2 show the extent to which the orbital motion contributes to the magnetization. On the other hand, for example for superconductors, $g' = 1$, showing that the diamagnetic currents which cause the Meissner effect are caused by electrons (see MEISSNER EFFECT; SUPERCONDUCTIVITY). For discussion of the measurement of the magnetomechanical ratio, see GYROMAGNFTIC EFFECT.

Spectroscopic splitting factor. This is the measure of the energy level splittings of an atomic system in a magnetic field. For a free electron spinning in a magnetic field H, the energy levels are split according to $\Delta E = g\mu_B H$ where μ_B is the Bohr magneton and g is the spectroscopic splitting factor, or g-factor, and is equal to 2.00. For electron spins in paramagnetic salts, there are complicated energy-level schemes because of the spin-orbit coupling and crystalline field interactions; the g-factor is different from case to case and may depend upon the orientation of the magnetic field with respect to the crystal axes. Under these circumstances, the g-factor is defined quantum mechanically. For ferromagnetic materials, the spectroscopic splitting factor is defined similarly; it is the factor in the ferromagnetic resonance condition which gives the splitting of the energy levels in the magnetic field. It is not generally equal to the magnetomechanical factor g', as the two are affected differently by the effects of orbital angular momentum. For free atoms, the spectroscopic splitting factor is identical to the Landé g-factor. For further discussion of g-factors, see MAGNETIC RESONANCE. [E.A.; F.K.]

Gyroscope

An instrument, often called simply a gyro, that maintains an angular reference direction by virtue of a rapidly spinning, heavy mass. Gyroscopes are used in airplanes to "remember" the orientation of the horizon and the direction of north during

maneuvers. Gyros measure the turning rate of gunsights to improve target tracking. The gyrocompass is a special north-seeking form of gyroscope developed to a high degree for shipboard use.

Gyros are used in missiles, airplanes, ships, and torpedoes as the basic element in automatic steering systems. A set of gyroscopes may be used to stabilize instrument platforms in maneuvering vehicles. There are stringent accuracy requirements for gyroscopes used for the stable platforms in inertial navigation systems.

Gyroscopic behavior is encountered in many natural phenomena, from the motions of atoms to the precession of planets.

Rigidity and precession. All gyroscope applications depend upon a special form of Newton's law which states that a massive, rapidly spinning body rigidly resists being disturbed and tends to react to a disturbing torque by precessing (rotating slowly) in a direction at right angles to the direction of the torque.

This principle can be readily demonstrated with the aid of a bicycle wheel which has been given a high spin velocity about its axle, as shown in Fig. 1*a*.

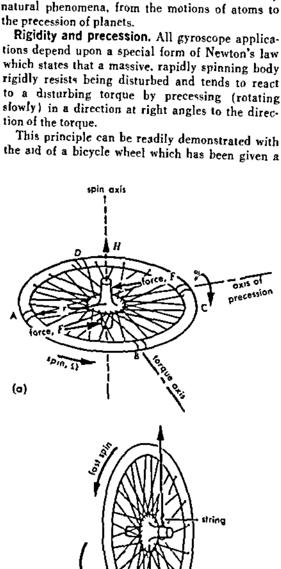


Fig. 1. Demonstrating the gyroscope principle. (*a*) Relation of spin, torque and precession axes. (*b*) Illustration of precession.

First, the property of rigidity can be demonstrated by supporting the axle vertically at one end and striking the other end a sudden blow. The axle appears not to be noticeably displaced, even under a heavy blow.

Next, to observe precession of the spinning wheel, the forces F are applied steadily. The wheel is found to precess slowly, not about the axis of the applied torque, but instead about an axis perpendicular to it (and also perpendicular to the spin axle). A variation of this experiment is shown in Fig. 1*b*. The weight of the bicycle wheel is supported by a string tied to one end of the axle so that gravity applies a torque to the wheel If the wheel is not spinning, it falls under gravity. But if the wheel is spinning rapidly it is observed not to fall. Instead the axle remains horizontal and precesses slowly about the string.

Gyroscope theory. The fact that the gyro precesses about an axis at right angles to the applied torque can be understood by focusing attention on a small piece of the rim as it spins around. Consider a piece that is initially at point A in Fig 1*a* At this location it feels the full force of the applied torque, and by Newton's law begins to accelerate downward. It does not acquire velocity immediately, however, but reaches its maximum downward velocity at point B. After passing point B the effect of the torque is to accelerate the particle upward so that, by symmetry, it has zero velocity again as it passes point C. Every particle experiences the same force pattern; the net effect is therefore a downward velocity of the rim at point B, an upward velocity at point D, and no vertical velocity at points A and C. The total result is a precession of the rim about axis AC proportional to the applied torque.

It is convenient to define a special property, known as angular momentum, of a spinning mass For the bicycle wheel of Fig. 1, the angular momentum H is

$$H = mr^2\Omega$$

where m is the mass of the rim, r is the radius of the rim, and Ω is spin velocity. The angular momentum embodies the gyroscope's rigidity, that is, its resistance to torque. For increasing values of H, the rate of precession ω decreases for a given torque T.

$$\omega = T/H \tag{1}$$

The axis of T is perpendicular to the axes of ω and of H (or Ω). This is the fundamental principle of a gyroscope. (The perpendicularity is sometimes indicated by the vector form $\mathbf{T} = \omega \times \mathbf{H}$.)

The spinning rotor need not be a simple rim, as it is in Fig. 1. The above principle applies to any spinning rotor that is perfectly symmetrical with respect to the spin axis. Any symmetrical rotor can be thought of as consisting of an aggregate of

rims, and the quantity H can be obtained by adding up the mr^2 values for all of the rims.

Nutation. The axle of a free gyro may be made to describe small circular or elliptical motions by applying an appropriate sudden torque impulse. The frequency f of such motions can be shown to be

$$f = \frac{1}{2\pi}\sqrt{H^2/I_1I_2}$$

in which H is angular momentum and I_1 and I_2 are moments of inertia of the assembly about each wheel diameter.

For practical gyroscopic instruments under reasonable disturbances, such motions are of very high frequency and are extremely small, so as to be of no importance. It is difficult to produce noticeable circular motions in the bicycle experiment of Fig. 1, for example. However, for academic investigation, special gyro configurations can be devised using slow spin velocity for which sizable circular motions are easily induced and observed. When they are superimposed on steady precession (due to unbalance) the result is a nodding up and down of the axle as it precesses about the vertical, which has led to the term nutation.

Gyroscopic instruments. The gyroscopic principle is used in instruments in several ways. Two are illustrated in Fig. 2, which shows instruments mounted to the structure of an airplane The free gyro uses the property of gyroscopic rigidity to sense changes in attitude; the single-axis rate gyro uses the precession characteristic to sense angular velocity.

In the free gyro of Fig. 2 the spinning wheel is isolated from the airplane by gimbals. When the airplane banks, the gyro remains vertical and furnishes the pilot with an artificial horizon reference. *See* AUTOMATIC HORIZON.

The rate gyro in Fig. 2 is one of a class of instruments known as single-axis gyroscopes. It is mounted to the aircraft structure by only a single gimbal so that if the airplane rolls (about its fore-and-aft axis) the rate gyro is obliged to roll with it. This causes the gimbal to turn (precess) against the stiff spring by a small angle which indicates rate of roll. A set of three such gyros, mounted with input axes mutually orthogonal, will serve to determine completely the total angular velocity of the aircraft in space.

Families of gyroscopic instruments are based on the free or on the single-axis arrangement.

Free gyroscopes. When a gyro is suspended so that its spin axis is not constrained to follow motions of its case, it is called free. The rotor may be supported on a single point (like a top) or by a set of gimbals (as in Fig. 2), or it may be "floated" in a liquid, on high-pressure air, or in a magnetic or electrostatic field. The rotor is usually spun relative to its suspension system by an electric motor, although in early gyros turbine buckets were cut in the rotor and driven by air jets.

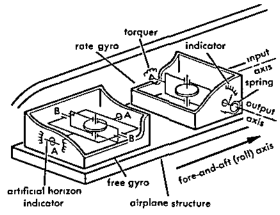


Fig 2 Two gyroscopic instruments When the airplane rolls, the free gyro remains fixed, the rate gyro precesses.

Two of the three angles defining the attitude of a vehicle can be measured by one free gyro. (The gyro cannot sense displacement about its spin axis.)

The performance of a free gyro is characterized simply by its fixity, or reciprocally stated by its random drift which depends upon its response to disturbing torques and upon the degree to which such torques can be eliminated Drift is discussed below under accuracy.

Vertical gyro This instrument, of which the artificial horizon of Fig. 2 is an early and still common application, has its spin axis maintained locally vertical It is used in aircraft to measure both bank angle and pitch attitude (the angle by which the nose is raised above the horizon).

In Fig. 2 the spinning rotor of the free gyro is supported by a set of two frames, the gimbals, which are arranged perpendicular to one another. The bearings AA, with which the outer gimbal is mounted to the airplane structure, and BB, with which the inner gimbal is mounted to the outer gimbal, are made to have the smallest possible friction. Then, because of its high angular momentum, the rotor of the free gyro is supposed to remain fixed in space as the airplane maneuvers about it. Thus if the airplane banks, the outer gimbal of the gyro will remain horizontal and can be used by the pilot as an artificial horizon as shown in Fig. 2. Alternatively, the angle between gimbal and case can be used to generate an electrical bank-angle signal for use in automatic pilot equipment (*see* AUTOPILOT). Pitch attitude is measured as the relative rotation between the inner and outer gimbals. The electrical signals are generated by precision-wound induction coils. *See* SERVOMECHANISM

To minimize cumulative drift, vertical gyros may be slaved, or caged, to the gravity vertical. Caging is automatically discontinued whenever the airplane is in a steady turn.

Directional gyro. The free gyro in Fig. 2 can be converted to a directional reference by remounting

its case with the spin axis horizontal and the indicator axis vertical. The indicator will then give the airplane's heading. Gyro drift is countered by slaving the instrument to the Earth's magnetic field, via either a separate magnetic compass or a flux valve device. See AIRCRAFT COMPASS SYSTEM.

Gyrocompass. This instrument seeks true north by aligning itself to the Earth's spin axis. A pendulous control maintains the gyro's spin axis horizontal, that is, perpendicular to a local vertical line, so that as the local vertical rotates with the Earth the instrument experiences a gyroscopic torque until its spin axis is parallel to the Earth's. See GYROCOMPASS

Apparent precession. A free gyro that is not slaved to any Earth reference (as the vertical and directional gyros and the gyrocompass are) will tend to maintain a fixed direction in inertial space, that is, with respect to fixed stars. Then as the

to precess.

Single-axis gyroscopes. A single-axis gyro, such as the rate gyro in Fig. 2, is suspended in just one gimbal whose bearings form its output axis. When the case of the instrument is mounted to a vehicle, motions of the vehicle about the instrument's input axis (perpendicular to both the spin and output axes) produce precession of the gyro gimbal within the output-axis bearings Gimbal motion is measured electrically.

Rate gyros. In a rate gyro the gimbal is restrained by a spring, so that its output angle θ is proportional to vehicle angular velocity. Thus, if the airplane in Fig. 2 has a constant rate of roll ω, then the principal torques about the output axis are the gyroscopic torque $H\omega$ and the spring torque $K\theta$, where K is the spring constant. Thus the instrument equation is:

$$\theta = H/K\omega \qquad (2)$$

The principal indices of performance of such an instrument are its sensitivity, accuracy, dynamic response, and drift. Sensitivity is the magnitude of the output angle per unit roll rate. From Eq. (2), this depends on the stiffness of the spring and the angular momentum. The accuracy with which the instrument senses a given rate depends upon the constancy of the spring stiffness and the consistency of the spring's null, or unstretched, position.

The dynamic response of the instrument depends again upon the spring stiffness and also upon the inertia of the wheel. A stiff spring leads to fast dynamic response at the expense of sensitivity, so that a design compromise must be made for each application. Sometimes the spring is replaced by an electrical feedback loop and torque motor, or torquer, which applies to the gimbal a restoring torque proportional to the gimbal angle. Dynamic response must be controlled by furnishing viscous damping between the gimbal and the case, so that transient oscillations are promptly damped out.

Drift of a single-axis gyro occurs when some unwanted torque appears about its output axis. For example, if the center of gravity of the rotor shifts, the gimbal will rotate and indicate an airplane rate when one exists. Gyro drift is discussed under a later section on accuracy.

Integrating gyros. For certain applications the spring can be removed from the rate gyro in Fig 2, and the instrument, now known as an integrating gyro, behaves somewhat differently.

For example, with strong viscous damping, the important torques about the gimbal output axis are the gyroscopic torque $H\omega$ and viscous damping torque, which is proportional to gimbal output-angle rate. Thus the output-angle rate of the gimbal is proportional to input angular velocity, and for ... is pro-

... put

... ing

... an

rate gyro. For small motions ... attitude signal similar to that obtained from the free gyro in Fig. 2. Integrating rate gyros are often employed instead of free gyros to control stable platforms because a single gyro gimbal can be made with greater precision than a pair of gimbals.

Yet another form of single-axis gyro is obtained if nonviscous fluid is used. Then the acceleration of the gimbal about the output axis becomes proportional to angular rate of the case about the input axis, so that if the case is tilted at a constant angle about the input axis then the gimbal rotates at a constant rate about the output axis. Such an instrument, known as a position-integrating gyro, is also used extensively to control stable platforms. It has some important advantages over the rate-integrat

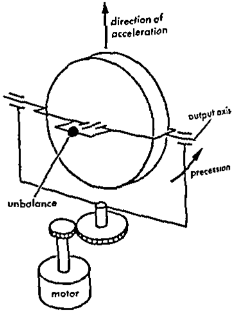


Fig. 3. Gyro velocity meter.

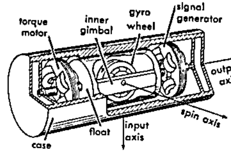


Fig. 4. Flotation of gimbal and rotor. (From *J. E. Gibson and F. B. Tuteur, Control System Components, McGraw-Hill, 1958*).

ing gyro, but requires more sophisticated electronic control of the stable platform.

Gyro velocity meter. Figure 3 depicts an ingenious method of using the gyroscopic principle to indicate vehicle velocity with great precision. The instrument is called a pendulous integrating-gyro accelerometer. The basic element is a single-axis gyro whose gimbal is given a deliberate unbalance. When the instrument is accelerated in the direction shown, an unbalance torque is created about the output axis. This torque is balanced by a gyroscopic torque created by precessing the gyro at just the proper rate with the motor. Then motor speed is just proportional to acceleration, and the number of revolutions the motor has made is just proportional to velocity. Such a device may be made sufficiently accurate for inertial guidance purposes.

Accuracy of gyroscopic instruments. Gyroscopes, both free and single-axis, may be classified by random drift rates into three broad groups. Gyros used to stabilize short-range missiles, torpedoes, and fire-control antennas have random drift rates measured in degrees per minute. The free gyros used for attitude reference in aircraft have typical unsupervised drifts measured in degrees per hour. The gyros used to stabilize inertial navigation equipment must be made to have drifts measured in degrees per year.

Gyro drift can be controlled only by extremely careful balancing, by reducing bearing friction to the smallest possible value, and by preventing temperature and magnetic effects from producing output axis torques. Control of these effects has been the subject of extensive fundamental research.

An important factor of the last class of gyros is the prevalent use of the flotation technique, whereby the friction level in the gimbal bearings is greatly reduced. The inner gimbal is a sealed float which houses the rotor. The float is suspended in a fluid whose density is just sufficient to relieve the gimbal bearings of the rotor weight; therefore, friction at these bearings is essentially zero. This construction is illustrated in Fig. 4.

Gyro-stabilized platform. A stable platform is an apparatus used to maintain angular reference directions in inertial space. The most effective application of a stable platform is made by inertial navigation systems which depend upon the platform for fixity to an extreme degree. In principle, a platform drift of 1° of arc can lead to a navigational error of 60 nautical miles *See* INERTIAL GUIDANCE SYSTEM.

To explain the stable platform, consider first the single-axis platform of Fig. 5*a*, in which the dip sensor may be either a free gyro or a single-axis gyro. In either case, any drift of the platform from the attitude prescribed by the gyro generates a feedback signal to the servo, which applies a corrective torque to the platform to drive it back where it belongs.

Use of the principle to control a platform about three axes is shown in Fig. 5*b* (in which integrating

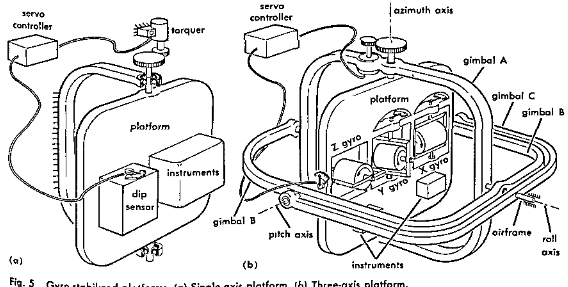


Fig. 5 Gyro-stabilized platforms. (a) Single-axis platform. (b) Three-axis platform.

single-axis gyroscopes happen to be shown). The servo loop of Fig. 5a is shown again in Fig. 5b. In addition, a signal from the X gyro would be used to apply restoring torque between gimbals A and B, and a signal from the Y gyro would be used to apply torque between gimbals B and C.

An extra gimbal (C) is shown in Fig. 5b. The purpose of this gimbal is to make it possible to mount the other three gimbals mutually orthogonal *to avoid gimbal torques. The redundant gimbal is* controlled by an electrical signal generated between gimbals A and B.

A platform like that in Fig. 5b can be made to have great rigidity in the face of the sometimes formidable disturbances imposed upon it by an aircraft environment, such as the rapid motions of the craft and the severe vibration and temperature extremes. At the same time ingenious design techniques are employed to minimize platform size and weight. [RBC]

Hackberry

A medium-sized to large tree, *Celtis occidentalis,* occasionally growing to a height of 120 ft It occurs in the eastern half of the United States, except in the extreme south, and is characterized by corky or warty bark, alternate, long-pointed, serrate leaves unequal at the base, and by a small drupaceous fruit, with thin, sweet, edible flesh. The pith of the twigs is chambered. The wood is used for furniture, boxes, and baskets. It is a shade tree and is also used for shelter belts.

Common hackberry, *Celtis occidentalis* (A H Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

Sugarberry, *Celtis laevigata,* is similar to hackberry. It grows in the southeastern United States and has narrower leaves with entire margins and smaller fruit. It is used for furniture, boxes, and baskets, for shelter belts, and as a shade tree *See* FOREST AND FORESTRY; TREE.

[A H G.]

Haemosporidiida

A relatively small and generally rather compact group of Sporozoa. The group is assigned the taxonomic status of a family by some authorities and the rank of a subclass by others. Most protozoologists probably still regard the Haemosporidiida as an order of the Telosporidia, a subclass of the Sporozoa. They are common protozoan parasites of vertebrates, and some of them are important causes of illness and death. The best known of the Haemosporidiida are the four species of malaria parasites of man. Not so well known is the existence of at least 50 other species of malaria, with a wide host distribution among terrestrial vertebrates, as well as numerous species of *Haemoproteus* and *Leucocytozoon*. The latter two genera are closely related to the genus *Plasmodium* in which all the true malaria parasites are placed.

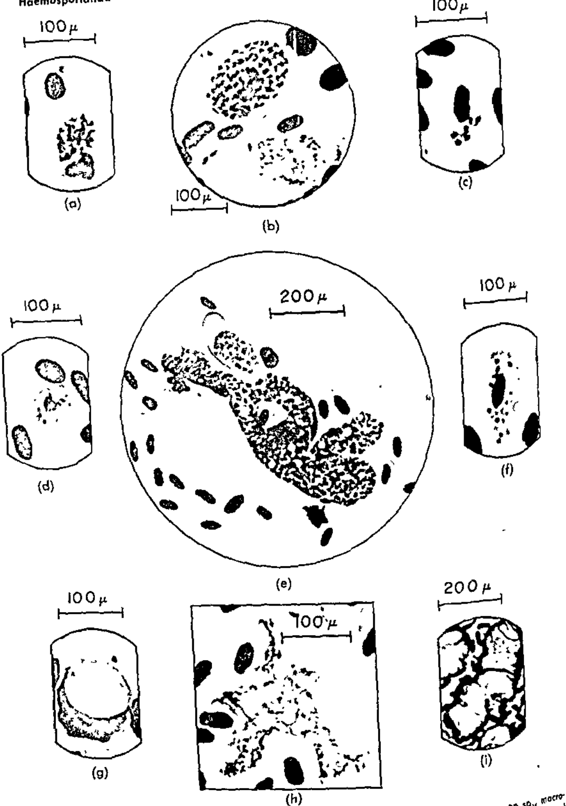
Transmission of these parasites is probably always effected by the bite of some blood-sucking invertebrate In the vertebrate host they reproduce asexually. Sometimes this occurs in the tissues of certain internal organs such as the lungs, liver, spleen, and brain; sometimes in the red blood cells, erythrocytes, or even in other types of blood and blood forming cells; and sometimes in both. The immature sex cells, gametocytes, occur in blood cells. These may be either erythrocytes or the white blood cells, leukocytes. The gametocytes mature into gametes after being ingested by the intermediate host. Fertilization ensues, with a subsequent period of development culminating in the production of numerous sporozoites. These forms are infective for the vertebrate host. Since these stages can develop no further in the arthropod, infection in this host is self limited. As a consequence, the parasite seldom does harm to the vector.

Taxonomy. Protozoologists disagree about the number of families that should comprise the order Haemosporidiida. Some think only the Haemoproteidae and Plasmodiidae should be included and there is good biological justification for combining these two. Others believe that a third family, the Babesiidae, also belongs in the order. Here, all three are included, though too little is known of their biological relationships to state with certainty that this arrangement is correct.

1. Plasmodiidae are Haemosporidiida inhabiting the erythrocytes of the vertebrate host (and, in a few species, other blood and blood-forming cells) in which both asexual reproduction and the formation of gametocytes occurs. Hemoglobin [illegible] lism results in the formation of a by-[illegible] hemozoin, typical of these parasites. The [illegible] brate host is probably always a mosquito.

2. Haemoproteidae are Haemosporidiida [illegible] which only the gametocytes occur in blood [illegible] asexual stages being confined to the inter[illegible] If in the erythrocytes, the parasites produce [illegible] zoin as in malaria. The invertebrate [illegible] known, is a blood-sucking fly.

3. Babesiidae are Haemosporidiida [illegible] blood cells (usually erythrocytes, but [illegible] is said, the lymphocytes) without [illegible] pigment. Life cycles are poorly known[illegible] sion is by ticks.



Haemosporidiida. (a) *Plasmodium gallinaceum*, segmenter. (b) *Plasmodium circumflexum*, exoerythrocytic segmenter. Lung of canary. (c) *Plasmodium hexamerium*. (d) *Plasmodium relictum*, macrogametocyte. Small black dots are hemozoin. (e) *Haemoproteus* sp., lung of song sparrow. (f) *Haemoproteus* sp., macrogametocyte almost encircling erythrocyte nucleus. Black dots are hemozoin granules. (g) *Leucocytozoon* sp., macrogametocyte from purple grackle. Enlargement and distortion of host cell, probably a lymphocyte. (h) *Leucocytozoon* sp., reproducing form in kidney. (i) *Leucocytozoon* sp., numerous immature gametocytes in kidney.

The Plasmodiidae are the true malarial parasites. Although the host spectrum is a rather broad one, the majority of known species are found in birds and reptiles, particularly lizards. The reptilian and avian species of *Plasmodium* parallel one another very closely. This suggests that parasitism in this group of Haemosporidiida may have occurred in birds and reptiles much earlier than in mammals where primates seem to be the preferred hosts. Malaria, however, also occurs in a number of other mammals such as rats, squirrels, bats, antelope, and water buffalo. Some mammalian species of *Plasmodium* have been transferred to the new genus *Hepatocystis*, because they reproduce only in tissues, as does *Haemoproteus*.

The vectors for many species of both genera still remain undiscovered. It is believed that species with mammalian hosts are usually transmitted by mosquitoes of the genus *Anopheles*, and those of birds by *Aedes* or *Culex*. This aspect of the life cycle of the reptilian malarias is unknown. Since well over 1600 species of mosquitoes exist, with almost equal variety in host preferences and biting habits, there is still ample room for such research.

Some of the better-known species of malarial parasites

Species of *Plasmodium*	Host	Vector
P. vivax	Man	*Anopheles* sp
P. malariae	Man (and probably chimpanzee)	*Anopheles* sp
P. falciparum	Man	*Anopheles* sp
P. ovale	Man	*Anopheles* sp
P. knowlesi	Monkey (especially *Macacus irus*)	Several species of *Anopheles* in laboratory, uncertain in nature
P. berghei	Congo tree rat, and other rats and mice	*Anopheles dureni*
P. cynomolgi	Monkey	?
P. inui	Monkey	?
P. gallinaceum	Domestic fowl	*Aedes* sp
P. lophurae	Fireback pheasant (and, in laboratory, domestic fowl and ducks)	?
P. hexamerium	Numerous species of passerine birds	?
P. circumflexum	Numerous species of birds, mostly passerine	*Culex* sp., *Theobaldia* sp
P. nucleophilum	Catbird, and others	?
P. relictum	Numerous species of passerine and other birds	*Culex* sp; *Aedes* sp; *Theobaldia*, probably some anophelines
P. cathemerium	Numerous species of birds, mostly passerines	*Culex* sp; possibly also some anophelines
P. vaughani	Robin (very common); other thrushes, rarely in other passerines	?
P. elongatum	Numerous species of passerine birds, also others	*Culex* sp.
P. polare	Cliff swallow (outside N America, in others)	?
P. rouxi	English sparrows in Near East	*Culex* sp. in laboratory (in nature?)
P. mexicanum	Lizards	?

It is noteworthy that many individuals are often malaria-resistant, even when bitten by a mosquito species known to be a very effective malaria vector. As a possible explanation, C. Huff showed that susceptibility to infection on the part of *Culex pipiens*, a good vector of the avian malaria parasite *Plasmodium cathemerium*, behaved as a simple Mendelian recessive, thus being genetically controlled.

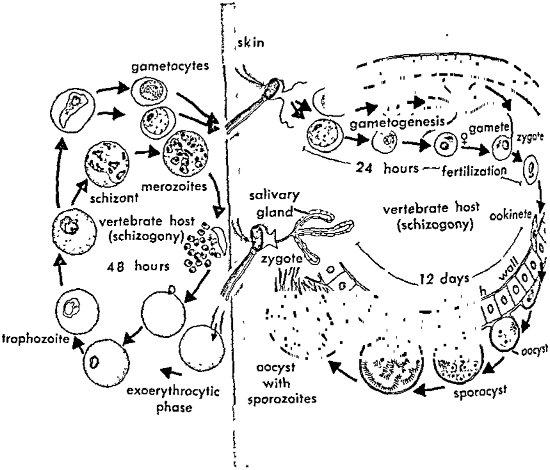
Included in the family Haemoproteidae are the two genera *Haemoproteus* and *Leucocytozoon*. It is believed (though without much proof) that the parasites of both genera are highly host-specific. If this is true, there are indeed many species. Infection by a member of one or both genera is very common in birds, especially among the passerines or perching forms. *Leucocytozoon* appears to be limited to avian hosts, but *Haemoproteus* also occurs in some reptiles.

Transmission of *Haemoproteus* seems usually to be by hippoboscid flies as in *Haemoproteus columbae* by *Pseudolynchia canariensis* (*maura*). *Leucocytozoon* has as its vector certain species of black flies (genus *Simulium*). The invertebrate host distribution in both genera may be much wider than this, for only in a very few species is the life cycle completely known.

The babesias (Babesiidae), causative agents of babesiasis, are the least known of the Haemosporidiida, even though the important cattle parasite *Babesia bigemina* was the first protozoan parasite for which an arthropod-vertebrate cycle was demonstrated (by T Smith and F. Kilbourne in 1893). Besides *Babesia* and *Theileria*, the genera *Toxoplasma*, *Dactylosoma*, *Anaplasma*, and *Nuttallia* are sometimes included in this family.

Pathogenicity. The Haemosporidiida vary greatly in pathogenicity. As with other large groups of parasites, while they are not, on the whole, very harmful to their hosts, there are some notable exceptions. Human malaria, though now on the wane in many parts of the world, was and is a very important human disease. The numerous species of *Plasmodium* occurring in the lower vertebrates may play a significant part in the maintenance of biological balances, particularly since such infection is often acquired early in the life of the host, when susceptibility is likely to be greatest.

Certain species of *Haemoproteus* and *Leucocytozoon* are known to be quite pathogenic to their avian hosts. Notorious examples are *Haemoproteus lophortyx* in various species of quail, *Leucocytozoon simondi* (*anatis*) in ducks, and *L. smithii* in turkeys. The incidence of these three species is



Diagrammatic life cycle of malaria. (*From T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957*)

use of insecticides. DDT is still one of the most valuable, despite the tendency of insects to develop resistance to it

Life cycles. The life cycles of the various Haemosporidiida (except those of the babesias, which are still very incompletely known) seem to be very similar, and are exemplified by the malaria parasites of man. Such life cycles also closely resemble those of the coccidia, a group of Sporozoa to which the Haemosporidiida are closely related, and from which they probably evolved.

Sporozoites (the infective forms) introduced by the vector develop into intracellular asexually reproducing stages (exoerythrocytic forms) in the tissues. In malaria, after several generations of this stage, some of the young parasites (merozoites) escape into the blood stream and infect red cells. Reproduction continues in these cells, sometimes for years, and may also continue in the body tissues. Some parasites go on to further reproduction, while others under unknown stimuli become gametocytes. *Haemoproteus* and *Leucocytozoon* undergo similar life cycles, except that reproduction is limited to the internal organs. The sexual phase of the cycle occurs only in the vectors, and involves gametogenesis, fertilization, growth of the oocyst, and eventual development of large numbers of sporozoites. Some of these migrate to the salivary glands, there to await an opportunity to reach another vertebrate host.

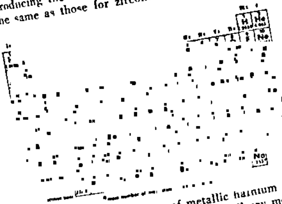
One of the interesting aspects of the haemosporidian life cycle is its frequent correlation with the physiological activities of the host. The periodicity of asexual reproduction is known to be conditioned by both the genetic constitution of the parasite and diurnal variations in host physiology, at least in some species. Relapses in *Leucocytozoon* infections and perhaps in malaria, are particularly common in the spring, due to seasonal changes in reproductive activity.

Biological aspects. Much of what is known about the basic biology of malaria is the result of research on the malarias of the lower animals. Earlier investigations were done chiefly on malaria infected birds, especially canaries, and, later, on ducks and chickens. *Plasmodium relictum* (*praecox*), *cathemerium*, *lophurae*, and *gallinaceum* were the species most used. Since 1950 *Plasmodium berghei*, a species occurring naturally in the African tree rat, has been found a very convenient tool for research on malaria in rats and mice. Their use has provided much knowledge about the life cycle of the malaria parasites, the mechanism of relapse, and chemotherapy. *See* CHEMOTHERAPY;

FERTILIZATION; GAMETOGENESIS; MALARIA; REPRODUCTION, ANIMAL; SPOROZOA. [R.D.M.]

Hafnium

A metallic element, symbol Hf, atomic number 72, atomic weight 178.5 (natural isotopes 174, 176, 177, 178, 179, 180). Hafnium is one of the less abundant elements in the earth's crust. It is found, mixed with zirconium, chiefly in such minerals as alvite, cyrtolite, and malacon. The methods for producing the metal from the ore are essentially the same as those for zirconium. *See* ZIRCONIUM.



The physical properties of metallic hafnium are similar to those of zirconium. It is a silvery metal that melts at 1975° and boils at about 5400°C. Its crystal structure is hexagonal close-packed below 1300°C and body-centered cubic above. Its density is 13.09 g/cm³, about twice that of zirconium, because its atomic weight is almost twice as great and its atomic volume about the same. Its cross section for absorption of neutrons is 120 barns (1 barn = 10^{-24} cm²), almost 700 times that of zirconium. The difference in neutron cross section makes it possible to analyze for hafnium in zirconium by neutron activation.

The chemistry of hafnium is almost identical with that of zirconium. The melting points, solubilities, and volatilities of the compounds of these two elements are very similar so that the information available for the compounds of one element can be assumed to apply to those of the other. The densities of hafnium compounds are greater than those of the comparable zirconium compounds. Thus the density of hafnia (HfO_2) is 9.68 g/cm³ at 20°C and that of zirconia (ZrO_2) is 5.73. This difference can be used to determine the ratio of the two elements in an oxide sample containing only those elements. The ratio of the two elements can also be determined spectrographically. The chemistry of this pair of elements is more similar than that of any other pair of elements because their ionic and atomic radii are almost equal and the arrangement of their valence electrons is the same. (Both are members of group IVb of the periodic table.)

The atomic radii of hafnium and zirconium are 1.57 and 1.58 A, respectively; the ionic radii, 0.84 and 0.87 A, respectively.

There is little commercial use of hafnium because of its limited supply and high cost. It has been used as filaments for lamps and as a getter in vacuum-tube manufacture. It may find use in the control of nuclear reactors because of its high neutron cross section. In the fall of 1958, the U.S. Bureau of Mines (Albany, Oreg.), announced pilot-plant production of 250-lb batches of high purity metal. *See* TRANSITION ELEMENTS. [D.D.CU.]

Bibliography: C. A. Hampel (ed.), *Rare Metals Handbook*, 1954; H. Remy et al., *Treatise on Inorganic Chemistry*, 1956.

Hail

Ice balls, occurring mainly in thunderstorms, composed usually of alternate layers of rime and hard ice, but varying between all hard ice (ice pellets) and all rime (soft hail or snow pellets). Small hail has a soft center and a single outer coat of hard ice. Hailstones are spherical or irregular, often conical; a typical diameter is ½ in. (1 cm), but sizes range to more than 5 in. in rare cases.

Hail in the United States is most frequent in a belt roughly from the southwest corner of North Dakota into the Texas Panhandle, and also from the southwest corner of Montana southward through western Utah. Such frozen precipitation is uncommon on the United States Pacific Coast and in Florida and is generally so throughout the tropics. Hail does extensive damage to crops and sometimes buildings, especially windows. Hailstorm paths vary greatly; a typical size is 20 miles long and 1 mile wide.

The meteorological conditions for true hail are (1) warm moist air in low levels, (2) thunderstorms with turbulence extending to great heights, sometimes over 50,000 ft, and (3) the freezing height roughly 10,000–12,000 ft above ground. Hail grows by alternate accumulation of rime and of liquid water which freezes to form hard ice. *See* CLOUD PHYSICS; PRECIPITATION (METEOROLOGY); SNOW; THUNDERSTORM. [J.R.F.]

Bibliography: *See* CLOUD PHYSICS.

Hair

An accessory structure of the integument of epidermal origin. Mammals are the only group of animals which possess true hair. At some stage during embryonic development, the body is covered with fine hair, the lanugo, which is usually shed before birth. The hair covering in land mammals serves the organism as an insulative structure, whereas marine forms such as the whale show an almost total lack of hair. Many animals shed their long thick coat of hair during the spring and there is a renewal of this coat in the fall. Baldness, in humans, seems to be sex hormone dependent, so that it behaves as a dominant character in males and recessive in females; not all types of baldness are hereditary nor are they sex-linked. Specialized types of hairs which are sensitive to stimuli, such as vibrissa, are found on the snout of many mammals. *See* HUMAN GENETICS; INTEGUMENT.

Development. In the basal layer of the epidermis, areas of extensive proliferation and downgrowth of cells are the clearly observable initial indications that follicles are forming. As a downgrowth into the dermis proceeds, a few dermal cells become concentrated and surrounded as the tip of the epithelial growth becomes concave. Later the concavity becomes egg-shaped with a narrow opening at the lower end, the dermal cells inside being the dermal papilla continuous through the opening with the connective tissue sheath which differentiates as the outer covering of the follicle.

The solid cord of epithelial cells above this bulb develops a bulge, to which the arrector pili muscle will attach, and a more distal bulge later which will become the sebaceous gland. Within the cord immediately above the region of this gland, keratinization occurs. When the central portion of this core falls out or is pushed out by the hair tip, the result is a hair canal to the surface lined with keratinized cells. *See* SEBACEOUS GLAND.

Sheath formation. Even before the bulb is fully formed and long before the follicle has reached its full length there is keratinization of the cells immediately distal to and descendant from the cells lining the distal or upper end of the concavity or incipient bulb. This represents the beginning of the internal sheath and appears just below the region [illegible]

[illegible] with a tapered upper end. Outside this sheath and inside the developing connective tissue sheath are the epithelial cells of the original cord, now called external sheath.

Hair formation. Hair formation begins shortly before the elongation of the follicle is complete. [illegible]

[illegible] around the relatively inert upper bulb, they keratinize, the outer ones forming the three layers of the inner sheath, the inner ones forming the cuticle of the hair, the cortex, and, especially in rodents, a distinct medulla. Both inner sheath and hair now grow distally, the former breaking down and sloughing as it reaches the region of the pilosebaceous canal. With the inner sheath, the hair tip breaks through the remnant of the original cord, the small region proximal to the hair canal. Pigment cells in the upper bulb, having entered passively with the epidermal downgrowth, become melanogenically active and deliver pigment granules to cells of the developing hair.

When hair growth ceases, the bulb everts from around the dermal papilla, the lower follicle degenerates, and a hair club forms. All that remains is the upper follicle to the point of muscle attachment, the condensed upper bulb region, which will serve as the germ for the new bulb, and a rounded clump of dermal papilla cells. In the new growth from this resting follicle there will again be elongation of follicle, bulb formation, and formation of internal sheath and hair. [R.B.C.]

Bibliography: H. B. Chase, Growth of the Hair, *Physiol. Revs.*, 34(1):113–126, 1954; W. Montagna and R. A. Ellis (eds.), *The Biology of Hair Growth*, 1958.

Hairworm

Any of several species of the class Nematomorpha of the p[illegible]

known [illegible]

hair sna[illegible]

blance to the hair from a horse tail, hence the superstition that horsehairs placed in a rain barrel will turn into these strange animals.

The different species of this group vary from ½ in. to 1 ft in length. They are all slender and cylindrical. Sexes are separate; the females are larger than the males. Young worms have a complete digestive tract which usually degenerates in the adult. There is no respiratory, circulatory, or excretory system. There is a single ventral nerve cord connected to a nerve ring around the esophagus.

The best known genus is *Gordius*. Adults are free-living in quiet water. The eggs are laid in surprisingly long strings, sometimes over 90 in. long and containing 6,000,000 eggs. Larvae have a brief free existence and penetrate almost any small aquatic animal that they contact. They can live and develop only in the body cavity of certain arthropods, notably grasshoppers, crickets, and beetles. Here they attain their full size after several weeks or months, and escape only when the host dies in or near water.

Early accounts indicated a two-stage parasitic life history, with the animal parasitic first in aquatic insects, then completing its development in land species. Such a cycle is now considered highly improbable. *See* NEMATOMORPHA. [J.D.B.]

Half-life

The time required for one-half of a given material to undergo chemical reactions; also, the average time interval required for one-half of any quantity of identical radioactive atoms to undergo radioactive decay.

The concept of the time required for all of the material to react is meaningless, because the reaction goes very slowly when only a small amount of the reacting material is left and theoretically an infinite time would be required. The time for half completion of the reaction is a definite and useful way of describing the rate of a reaction.

The specific rate constant k provides another way of describing the rate of a chemical reaction. In a first-order reaction

$$k = \frac{2.303}{t} \log \frac{c_0}{c}$$

where c_0 is the initial concentration and c is the concentration at time t. The relation between the

cific rate constant and period of half-life, $t_{1/2}$, in a first-order reaction is given by the equation

$$t_{1/2} = \frac{2.303}{k} \log \frac{1}{1/2} = \frac{0.693}{k}$$

In a first-order reaction, the period of half-life is independent of the initial concentration, but in a second-order reaction it does depend on the initial concentration according to the relation

$$t_{1/2} = \frac{1}{kc_0}$$

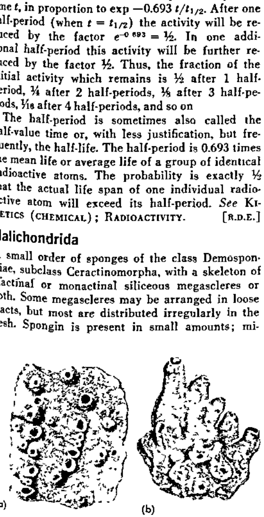
[F.D.]

Radioactive decay. The activity of a source of any single radioactive substance decreases to one-half in one half-period, because the activity is always proportional to the number of radioactive atoms present. For example, the half-period of Co^{60} (cobalt-60) is $t_{1/2} = 5.3$ yr. Then a Co^{60} source whose initial activity was 100 curies will decrease to 50 curies in 5.3 yr. The activity of any radioactive source decreases exponentially with time t, in proportion to $\exp -0.693\, t/t_{1/2}$. After one half-period (when $t = t_{1/2}$) the activity will be reduced by the factor $e^{-0.693} = \frac{1}{2}$. In one additional half-period this activity will be further reduced by the factor ½. Thus, the fraction of the initial activity which remains is ½ after 1 half-period, ¼ after 2 half-periods, ⅛ after 3 half-periods, $\frac{1}{16}$ after 4 half-periods, and so on

The half-period is sometimes also called the half-value time or, with less justification, but frequently, the half-life. The half-period is 0.693 times the mean life or average life of a group of identical radioactive atoms. The probability is exactly ½ that the actual life span of one individual radioactive atom will exceed its half-period. *See* KINETICS (CHEMICAL); RADIOACTIVITY. [R.D.E.]

Halichondrida

A small order of sponges of the class Demospongiae, subclass Ceractinomorpha, with a skeleton of diactinal or monactinal siliceous megascleres or both. Some megascleres may be arranged in loose tracts, but most are distributed irregularly in the flesh. Spongin is present in small amounts; microscleres are absent. A skinlike dermis is present and is often reinforced with tangentially placed spicules.



Halichondria panicea. (a) Encrusting form. (b) Fistular form. (After Bowerbank)

Halichondrid sponges are encrusting, massive, lobate, or branching in shape. Common shallow-water species, such as *Halichondria panicea*, exhibit extensive intraspecific variations in shape associated with environmental conditions. Halichondrids inhabit all seas, occurring chiefly in tidal areas and shallow waters of the continental shelf. Some species occur down to depths of at least 1500 meters. Fossil species are unknown. *See* DEMOSPONGIAE. [W.D.H.]

Halide

A compound of the type MX, where X may be fluorine, chlorine, bromine, or iodine, and M may be another element or organic radical. The compounds are considered derivatives of the hydrohalic acids, HX, where the halogen X has an oxidation state of -1. All the halides form both covalent and ionic inorganic salts, coordination compounds, and many organic compounds.

Many properties show a gradual change as one compares the halides. The size of the ions increases from F^- to I^-. The strength of the halide as a reducing agent also increases from F^- to I^-. It becomes more difficult to produce the element from the halide as one goes from I^- to F^-, electrolysis being required to produce elemental fluorine. The stability of complexes formed generally increases from I^- to F^-. The solubilities of metal halides in water increase going from I^- to F^-. As a group the halides are very soluble. *See* BROMIDE; CHORIDE; COMPLEX COMPOUNDS; FLUORIDE; HALOGEN ELEMENTS; HALOGENATED HYDROCARBON; HALOGENATION; IODIDE. [E.E.WR.]

Halimeda

A genus of green algae belonging to the family Codiaceae. The plant is small and bushy, composed of thick leaflike segments. These show considerable ... tion in size and shape in different species and ... seas. In favored areas, such as reef flats and the lagoons of atolls, *Halimeda* plants develop in enormous numbers to form *Halimeda* meadows. The genus is important as a fossil and as a limestone builder It has a known geologic range from Early Cretaceous to the present day. *See* ALGAE FOSSILS; ATOLL; ORGANIC REEF. [J.H.J.]

Bibliography: E. S. Barton, *The Genus Halimeda*, Siboga Expedition Monograph 60, 1901; W. R. Taylor, *Plants of Bikini and Other Northern Marshall Islands*, 1950.

Hall effect

The development of a transverse electric potential gradient in a current-carrying conductor placed in a magnetic field when the conductor is positioned

so that the direction of the magnetic field is perpendicular to the direction of current flow. Analysis of the Hall effect yields important information on the band structure of metals and semiconductors and on the nature of the conductivity process. In semiconductor research, the magnitude of the Hall effect in simple cases provides a direct estimate of the concentration of charge carriers. The Hall effect is one of the so-called galvanomagnetic effects. See GALVANOMAGNETIC EFFECTS.

The experimental arrangement for observing the Hall effect is shown schematically in Fig. 1. It is observed that a voltage V_H appears between the sides of the specimen whenever the current density J_x and the magnetic field B_z are nonvanishing. The electric field $E_H = V_H/d = E_y$, using the coordinates of Fig. 1, is found to be proportional to the product of J_x and B_z. The quantities V_H and E_H are called the Hall voltage and Hall field, respectively.

Physical interpretation. A simple interpretation of the Hall effect may be given by the following argument. Each charge carrier in the solid is assumed to move with a drift velocity $v_{dx} = J_x/ne$. Here n is the number of charge carriers per unit volume and e is the charge of each carrier. Each charge carrier experiences a Lorentz force

$$F_{yB} = -\frac{e}{c} v_{dx} B_z \qquad (1)$$

where c is the velocity of light. Unless this force is compensated by some other force, the charge carriers will acquire a drift velocity in the negative y direction, giving rise to a component of current in the y direction. The experimental arrangement, however, requires that $J_y = 0$. The necessary compensating force is provided by the Hall field E_y, whose magnitude and direction is such that

$$F_y = F_{yB} + F_{yE} = 0 = -\frac{e}{c} v_{dx} B_z + eE_y \qquad (2)$$

Hence

$$E_y = \frac{v_{dx} B_z}{c} = \frac{J_x B_z}{nec} \qquad (3)$$

The Hall coefficient R, defined by the ratio

$$R = E_y / J_x B_z \qquad (4)$$

is thus equal to

$$R = 1/nec \qquad (5)$$

A similar result can be obtained by solving the Boltzmann transport equation. The Hall coefficient of a metal provides the most direct information of both sign and number of charge carriers, provided the free electron model is valid (see FREE-ELECTRON THEORY OF METALS). The sign of R is the same as the sign of the charge e, and the magnitude of R is inversely proportional to n. The number of conduction electrons per unit volume calculated from measured Hall coefficients agrees reasonably well with the product of valence and number of atoms

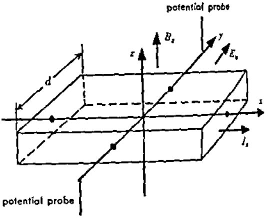


Fig. 1. Hall effect.

per unit volume for the monovalent metals. Erratic results are often obtained for polyvalent metals, as is apparent from the table.

In these cases a two-band model is frequently invoked in the interpretation of the Hall effect and other transport phenomena. Because n is almost independent of the temperature T, R also changes only slightly with temperature.

Thermal side effects. It can be shown that an isothermal condition can be maintained in a specimen only if a thermal current can flow freely in the transverse direction. Unless the sample is immersed in a constant temperature bath, $\partial T/\partial y \neq 0$. Under nonisothermal conditions, the finite temperature difference between the sides of the sample may give rise to spurious thermal emfs at the potential probes. The generation of a transverse temperature gradient by a current flowing in the presence of a transverse magnetic field is known as the Ettingshausen effect, and proper correction for this effect should be made. A convenient way of avoiding these thermal difficulties is to use an alternating current as the primary current J_x, thereby preventing the establishment of a significant temperature gradient in the sample.

Number of valence electrons and $n = 1/Rec$ for several metals*

Metal	$N \times 10^{-22}$	$NZ \times 10^{-22}$	$(1/Rec) \times 10^{-22}$
Monovalent metals			
Lithium	4.6	4.6	3.7
Sodium	2.5	2.5	2.5
Potassium	1.3	1.3	1.5
Cesium	0.85	0.85	0.8
Copper	8.5	8.5	11.4
Silver	5.8	5.8	7.5
Gold	5.9	5.9	8.7
Polyvalent metals			
Beryllium	12.3	24.6	-2.5
Magnesium	4.3	8.6	6.7
Zinc	6.6	13.2	-18
Cadmium	4.6	9.2	-10.5
Aluminum	6.0	18.0	21
Indium	3.8	11.4	89
Thallium	3.5	10.5	-26
Tin	3.7	14.8	156

* N is number of atoms per cm^3; Z is valence of atoms

Effect in ferromagnetic metals. Ferromagnetic metals exhibit anomalously large Hall effects at temperatures below the ferromagnetic Curie temperature. Moreover, the Hall voltages are not linear functions of B. Typical curves are shown in Fig. 2. The knees of the curves occur at saturation, and the shapes of the curves suggest that E_H be written in the form $E_H = R_0H + R_1M$. Here M is the magnetization, and R_0 and R_1 are designated the ordinary and extraordinary—or anomalous—Hall coefficients, respectively. The coefficient R_0 is not a sensitive function of temperature; R_1 changes by orders of magnitude as the temperature is increased from 4°K to the Curie temperature (Fig. 3)

Effect in semiconductors. In semiconductors the number of electrons (or holes) is generally sufficiently small so that the energy distribution approximates very closely the classical distribution. One can show that if the charge carriers are scattered predominantly by lattice vibrations

$$R = \frac{3\pi}{8} \cdot \frac{1}{nec} \qquad \text{(lattice scattering)} \qquad (6)$$

and if scattering by ionized impurities is the dominant mechanism

$$R = \frac{315\pi}{512} \cdot \frac{1}{nec} \qquad \text{(ionized impurity scattering)} \qquad (7)$$

As in metals, the sign and magnitude of R determine the sign and number of charge carriers, provided only one type of carrier (electrons or holes) is present. In semiconductors at high temperatures (intrinsic region) R decreases approximately exponentially with increasing temperature, largely because of the exponential increase of the number of charge carriers with increasing temperature. At low temperatures (extrinsic region) R is approximately constant. In the intrinsic region, and in some cases also in the extrinsic region, the Hall coefficient must be evaluated with the aid of a two-band model because more than one type of carrier is responsible for charge transport. *See* SEMICONDUCTOR.

Effect in ionic crystals. Alkali halides become electronic conductors only under the influence of light that will excite electrons from the valence band to the conduction band. Because the photoelectrons are few in number, they form a classical electron gas. Measurements of the Hall effect and of photoconductivity provide information on the mobility of these electrons. The scattering mechanisms and the effective masses of conduction electrons in ionic crystals are greatly influenced by the local polarization of the crystal by these charge carriers, and therefore the usual theories of mobility are not directly applicable here. For this reason, Hall effect and conductivity measurements in ionic crystals are of considerable interest. The measurements require rather special techniques and are more difficult than in metals or semiconductors. The data that are available indicate that current theories correctly describe the interaction between conduction electrons and lattice vibrations in alkali halides. [F.J.B.]

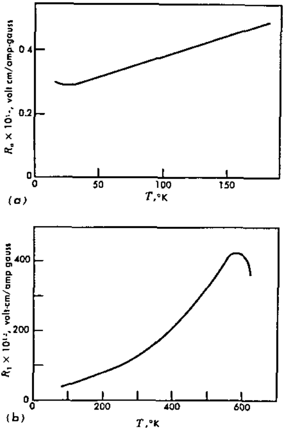


Fig. 3. Hall coefficients of nickel as a function of temperature: (a) ordinary coefficient (after J. P. Jan and H. M. Gijsman, *Physica*, vol. 18, 1952); (b) extraordinary coefficient (after J. P Jan, *Helv Phys Acta*, vol. 25, 1952)

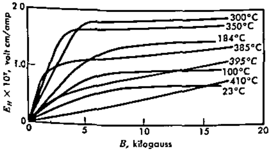


Fig. 2. The Hall field per unit current density in nickel as a function of magnetic induction between room temperature and 410°C. (After A. W. Smith, *Phys. Rev.*, vol. 30, no. 1, 1910)

Bibliography: N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals*, 2d ed., 1948; F. Seitz and D. Turnbull (eds.), *Solid State Physics*, vol. 1, 1955, vols. 4 and 5, 1957; E. Spenke, *Electronic Semiconductors*, 1958; A. H. Wilson, *The Theory of Metals*, 2d ed., 1953.

Halley's comet

A member of the solar system having an orbit with a period of about 76 years. The radius of the nucleus of Halley's comet has been estimated as 15 kilometers. Its large size and the well-situated orbit (with perihelion about half-way between the Sun and Earth's orbit) enable it to be seen by the naked eye each time it approaches perihelion. Its appearance has been recorded in the West, China, or Japan for all but one of the returns since 240 B.C.

Orbital elements. In 1705, Edmund Halley employed Newton's new gravitational theory in the computation of cometary orbits. Of the 24 comets studied, Halley observed that those of 1682, 1607, and 1531 had remarkably similar orbits. Correctly assuming that they were the same object, he predicted its return in 1758. As a tribute to the author of this work, the comet has borne his name since then.

The comet has a retrograde orbit. Aphelion lies between the orbits of Neptune and Pluto. The last perihelion passage was on April 19, 1910. The other five orbital elements (*see* CELESTIAL MECHANICS) are

a	e	i	ω	Ω
17.8 AU	0.967	162°	112°	57°

The position of the orbit in space is such that it approaches Earth's orbit to within 0.15 astronomical units (AU) at the ascending node and 0.05 AU at the descending node.

Associated meteors. Two meteor showers (*see* METEORS), the Orionids and the η-Aquarids, are observed each year at about the time Earth is at these points of its own orbit. The distance from the comet's orbit to Earth's orbit is considerably larger than in other cases of comet-meteor shower associations. However, the

from the comet and have since been perturbed by planets into their present orbits, which intersect that of Earth. *See* COMET. [R.E.MC.]

Bibliography: N. T. Brobnikov, Halley's comet in its apparition of 1910–11, *Lick Observatory*, vol. 17, 1931.

Halloysite

A group of clay minerals made up of the same structure units as kaolinite. The major use of halloysite clay is in the manufacture of cracking petroleum catalysts. Pure halloysite clays are not abundant, and hence their potential economic use has not been explored. *See* CLAY MINERALS; KAOLINITE.

There are two types of halloysite. Both are light-colored; one is porous and friable, while the other is dense, nonporous, and porcelainlike. The porous variety has the same chemical composition as kaolinite, whereas the nonporous hydrated form contains an added amount of water ($2H_2O$). The hydrous variety has a larger c-axis spacing than kaolinite and changes upon dehydration to about the same dimension as kaolinite. The transition from the higher hydrated form to the lower hydrated form begins at temperatures as low as 60°C and is not reversible. Temperatures of the order of 400°C are necessary for complete removal of this interlayer water. The nomenclature of the types of halloysite is confusing.

Structure and morphology. The halloysite minerals are made up of the same structure units as kaolinite, that is, layers composed of single silica tetrahedral and alumina octahedral units. They differ in the presence of a single layer of water molecules between each silicate layer in the hydrated form. The basal spacing of the dehydrated form is about 7.2 angstrom units (A) or about the thickness of the kaolinite layer, and the basal spacing of the hydrated form is about 10.1 A. The difference, 2.9 A, is about the thickness of a single molecular sheet of water molecules.

Halloysite also differs from kaolinite in that there is random displacement of silicate layers in both the a and b crystallographic directions.

Electron micrographs show that halloysite is composed of elongate particles which appear to be tubes (see figure). The tubes of the hydrated $4H_2O$ form appear to be made up of overlapping curled-up sheets of the kaolinite type. The presence of interlayer water molecules and the irregular stacking of units would facilitate this curvature. On dehydration the tubes frequently collapse, split, or unroll.

Other characteristics. The cation-exchange capacity of the lower hydration form of halloysite is 5–10 milliequivalents per 100 grams, whereas that

Electron micrograph of halloysite, British Guiana. (From R. E. Grim, Clay Mineralogy, McGraw-Hill, 1953)

of the hydrated variety is somewhat higher. The major cause of this exchange capacity is broken bonds.

Differential thermal curves for halloysite show a loss of pore water and interlayer water at about 100°C. Above about 200°C the dehydration characteristics are essentially like those for kaolinite. The high-temperature phases formed on firing halloysite are like those for kaolinite, but there are some differences in their development. Thus halloysite does not uniformly vitrify and its fusion point is higher than that of kaolinite.

Halloysite is often found in hydrothermal deposits. It has also been reported in weathering products; however, it is a rare component of such materials. It is generally absent in sedimentary rocks, and it has been suggested that metamorphic processes would change halloysite to kaolinite. [F.M.W.; R E.GR.]

Hallucination

A fixed and incorrigible report of a perception without an appropriate stimulus. Frequently the stimulus is placed into the outside world, although it can be demonstrated that it is within the person. At times it may be difficult to differentiate hallucinations from fantasies. Both can be concrete or abstract, lively or shadowy, realistic or bizarre; fantasies are usually recognized as such and as being different from perceptions and also from hallucinations, which are often mistaken for true perception.

Organic disturbances may cause hallucinations, although in most instances they can be explained by psychological and cultural circumstances. An individual's drives and emotions largely govern his hallucinations, which are primitive experiences and usually have the earmark of the person's cultural background. It is well known that a thirsting person in desert country will have hallucinations of wells and oases. Once visions of saints and devils were frequent; today they are rare, having been replaced by hallucinations of the devices of the machine and electronic age.

Hallucinations have the characteristics of one or several sensory systems; they may be visual, auditory, tactile, olfactory, or gustatory. In adults in the western world, hallucinations should raise the question of mental pathology. In primitive individuals and in primitive cultures in which myth and reality are not clearly differentiated, hallucinations occur more frequently. [F.C.R.]

Bibliography: N Cameron and A. M. Cameron, *Behavior Pathology*, 1951.

Halo

A system of luminous rings or arcs around the sun or moon, caused by refraction or reflection of light rays illuminating ice crystals floating in the air. Since the ice crystals have the form of hexagonal pyramids or prisms, the rays of minimum deviation in the prismatic refraction on the sides and the top or bottom of these prisms form a cone centered around the direction toward the sun. These rays are seen as luminous circles around the sun with a radius of either 22° (inner halo) or 46° (outer halo). Because the minimum deviation of the emerging rays is smaller in the red than in the blue, the inner and outer halo show a distinct red tint on the boundary toward the sun. Besides the inner and outer halo several similar arcs, luminous arcs, and bright spots (*see* SUNDOG) are occasionally seen in the sky as a result of refraction or reflection of light rays illuminating ice crystals in the atmosphere. All these phenomena are usually considered parts of a more complex system, called the halo. The inner halo is, however, most frequently observed. [Z.S.]

Halocline

A layer of sea water within which a marked salinity gradient is present. A permanent halocline of decreasing salinity usually coincides with the top of the permanent thermocline. Although weak and generally less important than the thermocline, the oceanic halocline furnishes significant clues to the origin and movement of water masses. Haloclines are extremely important in estuarine oceanography, however, as their depth, strength, and shape greatly influence circulation, mixing, and flushing of these embayments. Estuarine haloclines have strong positive gradients which are formed as fresh water flows seaward over sea water entering along the bottom. *See* ESTUARINE OCEANOGRAPHY; SEA WATER, THERMOCLINE. [J.J.SC.]

Haloform reaction

The halogenation of acetaldehyde or a methyl ketone in aqueous basic solution. The reaction is characteristic of compounds that contain a CH_3CO— group linked to a hydrogen or to another carbon atom. The halogen is dissolved in aqueous base (sodium hydroxide or carbonate) to form an alkali hypohalite, which is the halogenating agent. This is added to the ketone. The reaction occurs in two phases: the methyl group is fully substituted by halogen to give an intermediate trihaloketone, Eq. (1); the trihaloketone breaks down in the presence of alkali to produce the haloform and the sodium salt of a monocarboxylic acid, Eq. (2).

$$RCOCH_3 + 3NaOX \rightarrow RCOCX_3 + 3NaOH \quad (1)$$
$$RCOCX_3 + NaOH \rightarrow RCOONa + HCX_3 \quad (2)$$

Sufficient alkali is produced in the initial halogenation to cleave the trihaloketone so that the reactions proceed simultaneously. The haloforms which can be produced by this method are chloroform, bromoform, and iodoform.

Since the solution of hypohalite is mildly oxidizing, compounds which are easily oxidized to acetaldehyde or to a methyl ketone also give the haloform reaction. Examples are ethyl, isopropyl, and *sec*-butyl alcohols. Curiously, the compound CF_3-$CCl{=}CCl_2$, on oxidation, is routinely cleaved by potassium permanganate and sodium hydroxide

to yield sodium and potassium salts of trifluoroacetic acid, without undergoing the haloform reaction. This attests to the stability of the CF_3— group *See* TRIFLUOROACETATE.

Acetone or ethyl alcohol reacts with calcium hypochlorite to produce chloroform. *See* CHLOROFORM; HALOGENATED HYDROCARBON. Sodium hypobromite and sodium hypoiodite react in the same manner to yield bromoform and iodoform, respectively.

The haloform reaction is so characteristic that it may be used as a dependable qualitative test in distinguishing methyl ketones and methyl carbinols. Iodoform, for example, is an easily identifiable halide which precipitates from an aqueous solution as a yellow solid. Similarly, the test may be used to distinguish between ethyl and methyl alcohols, since ethyl alcohol forms a precipitate of iodoform whereas methyl alcohol does not. All methyl ketones, however, give the iodoform test. Secondary alcohols which can be oxidized to methyl ketones likewise undergo the haloform reaction. *See* IODOFORM.

In addition to use as a test reaction and as a method of synthesis for haloforms, the haloform reaction may be used to prepare acids from methyl ketones when conventional methods of synthesis are not feasible. Treatment of the ketone with a basic solution of iodine yields iodoform together with a solution of the acid salt. After filtration of the iodoform, and subsequent acidification of the basic solution, the acid is obtained routinely.

The haloform reaction was formerly applied industrially, specifically in the manufacture of chloroform, but it has been virtually displaced by more economical methods. [E.CL.; M.C.G.]

Halogen elements

The halogen family consists of the elements fluorine, F; chlorine, Cl; bromine, Br; iodine, I; and astatine, At. The relative abundances of these elements in seawater in parts per million (ppm), and in the lithosphere (earth's crust to a depth of 10 miles) has been estimated as follows.

	Seawater	Lithosphere
Fluorine	1.4 ppm	770 ppm
Chlorine	18,980 ppm	550 ppm
Bromine	65 ppm	1.6 ppm
Iodine	0.05 ppm	0.3 ppm

Astatine, previously known as alabamine, is a radioactive element whose longest-lived isotope, At^{210}, has a half-life of only 8.3 hours. It can occur naturally only through the continuous decay of elements such as U^{235} and Th^{232}, and the total astatine present at any one time in the outer mile of the earth's crust has been estimated at 6.86 milligrams. This is primarily At^{215}, with a half-life measured in seconds. The longer-lived isotopes are generated only in the laboratory. Measurements of the chemical and physical properties of astatine are made in very dilute solutions with difficulty and with limited accuracy so that its properties cannot yet be generalized with those of the other halogen elements.

The halogens are the best-defined family of elements. They have an almost perfect gradation of physical properties. The increase in atomic weight from fluorine through iodine is paralleled by increases in density, melting and boiling points, critical temperature and pressure, heats of fusion and vaporization, and even in progressively deeper color (fluorine is pale yellow; chlorine, yellow green; bromine, dark red; and iodine, deep violet). The atoms of each have seven electrons in their outer shells and normally have a negative valence of 1 in compounds. They exist individually as diatomic molecules, such as Cl_2, although several of the many possible interhalogen compounds are known, of the type XX'_n, where $n = 1, 3, 5$, or 7.

Chemically, fluorine is the most powerful oxidizing agent known. The heavier halogens have progressively less oxidizing ability. Each forms an acid with hydrogen, and salts with metals. The properties of these acids and salts show as consistent a relationship as the elements themselves. Organic halogen compounds generally show progressively increased stability in the order iodine, bromine, chlorine, fluorine.

Although all halogens generally undergo the same types of reactions, the extent and ease with which these reactions occur vary markedly. Fluorine in particular has the usual tendency of the lightest member of a family of elements to exhibit reactions not comparable to the other members. Each halogen must be considered individually, both in its preparation and in its reaction. *See* ASTATINE; BROMINE; CHLORINE; FLUORINE; HALIDE; HALOGENATION; IODINE; PERIODIC TABLE. [A.A.G.]

Halogen minerals

Naturally occurring compounds containing a halogen as the sole or principal anionic constituent. There are over 70 such minerals but only a few are common and can be grouped according to the following methods of formation.

1. Saline deposition by evaporation of sea water or salt lakes. Halite (rock salt), NaCl, is the most important of this type and is found in beds covering many hundreds of square miles and ranging in thickness from a few feet to over 1000 ft. Of the other minerals associated with halite, sylvite (KCl) and carnallite ($KMgCl_3 \cdot 6H_2O$) are the most important. *See* EVAPORITE (SALINE).

2. Hydrothermal deposition. Fluorite, CaF_2, is the chief representative of this type and occurs in veins by itself or associated with metallic ores. Cryolite, Na_3AlF_6, may be of primary deposition or result from the action of fluorine-bearing solutions on preexisting silicates.

3. Secondary alteration. Chlorides, iodides, or bromides of silver, copper, lead, or mercury may form as surface alterations of ore bodies carrying these metals. The most common are cerargyrite, AgCl, and atacamite, $Cu_2(OH)_3Cl$.

4. Deposition by sublimation. Halides formed as sublimation products about volcanic fumaroles include salammoniac, NH_4Cl; malysite, $FeCl_3$; and cotunnite, $PbCl_2$. At Mount Vesuvius, Italy, is the most noted occurrence of such minerals.

5. Meteorites. Lawrencite, $FeCl_2$, has been found in iron meteorites. *See* METEORITE.

See also CERARGYRITE; CRYOLITE; FLUORITE; HALIDE; ROCK SALT. [C.S.HU.]

Bibliography: C. W. Correns, The geochemistry of the halogens, in L. H. Ahrens, K. Rankama, S. K. Runcorn (eds.), *Physics and Chemistry of the Earth*, vol. 1, 1956.

Halogenated hydrocarbon

One of a group of halogen derivatives of organic hydrogen- and carbon-containing compounds. The group includes monohalogen compounds (alkyl or aryl halides) and polyhalogen compounds that contain the same or different halogen atoms. The halogenated hydrocarbons may be classified under the general headings of alkyl halides, unsaturated (vinyl and allyl) halides, polyhalogenated compounds, and halogenated aromatic hydrocarbons.

Uses. The alkyl halides, especially bromides and iodides, are used extensively as research intermediates. Several are essential pharmaceutical intermediates. Alkyl chlorides are important industrially in the manufacture of other organic compounds, especially alcohols and esters.

The polyhalogenated hydrocarbons find many uses. They are generally chemically inert, and are widely used in many industrial processes as solvents, degreasers, and dry-cleaning agents. The polybromides, more reactive than the polychlorides, are frequently used in the synthesis of other compounds. The polyfluorides, often with attached chlorine and bromine, are called freons and are widely used as refrigerants. The unsaturated polyhalogen compounds are also valuable in a variety of organic syntheses.

The halogenated aryl hydrocarbons are used in organic syntheses as solvents and as reactants. Generally, the chlorine compounds are more inert than the bromine and iodine compounds.

The aralkyl- or benzyl-type halides are more reactive than the aryl halides. They are used chiefly as chemical intermediates.

Alkyl halides. These are monohalogenated derivatives, RX, of the alkanes, or saturated hydrocarbons. Typical of these are methyl iodide (CH_3I), ethyl chloride (CH_3CH_2Cl), and *n*-propyl bromide ($CH_3CH_2CH_2Br$). The alkyl chlorides and bromides may be obtained by direct halogenation of the hydrocarbons. Thus, alkyl chlorides are prepared by direct chlorination of alkanes, particularly methane, ethane, propane, butane, and pentane. Chlorine gas is used alone, with light or heat to promote reaction, or with catalysts such as aluminum chloride, ferric chloride, and other metallic chlorides. At lower temperatures, the relative rates of formation of chlorides are in the order tertiary > secondary > primary. At higher temperatures, this tendency is less significant. Chlorination may proceed so vigorously that special techniques are necessary to avoid excessive decomposition that causes low yields of desired products. The process may be controlled frequently by using an excess of the hydrocarbon, an inert solvent, or a diluent inert gas. In all cases, mixtures of monochloro, dichloro, and polychloro compounds result; these can be separated, usually, by fractional distillation.

Direct bromination of the alkanes is restricted to the simple paraffins and olefins (unsaturated hydrocarbons). The reactions are generally conducted either in the vapor or in the liquid phase, and heat and light are used as initiators. Because oxygen decreases the rate of reaction (this indicates a free-radical mechanism), the reactions are run in an inert atmosphere. Higher temperatures favor substitutive bromination; on the other hand, bromine adds best to olefins at lower temperatures.

Iodine does not react with hydrocarbons. Fluorine reacts violently and causes extensive decomposition unless the reaction is carefully controlled.

Alkyl halides (chlorides, bromides, iodides, fluorides) may also be prepared by the addition of hydrogen halide to olefins. The order of reactivity is HI > HBr > HCl. With an unsymmetrical olefin, the addition of the hydrogen halide follows Markownikoff's rule. This rule states that the negative radical (F, Br, Cl, I) attaches to the carbon atom which has the smaller number of hydrogen atoms. Thus, propylene ($CH_3CH{=}CH_2$) adds hydrogen iodide to form isopropyl iodide (CH_3CHICH_3) and not *n*-propyl iodide ($CH_3CH_2CH_2I$), Eq. (1). Exceptions to Markownikoff's rule are those reactions involving the addition of hydrogen bromide. In the absence of oxygen, normal addition is effected. However, when peroxides and oxygen are present, the addition may be abnormal. For example, propylene, in the presence of peroxides and oxygen, adds hydrogen bromide to produce *n*-propyl bromide and not the expected isopropyl bromide, Eq. (2). This is explained on a basis of the difference in reaction mechanisms. Normal addition is ionic in character, whereas peroxide-catalyzed addition is a free-radical process. *See* ADDITION REACTION.

$$CH_3CH{=}CH_2 + HI \rightarrow CH_3CHICH_3 \qquad (1)$$

$$CH_3CH{=}CH_2 + HBr \xrightarrow[\text{Oxygen}]{\text{Peroxides}} CH_3CH_2CH_2Br \qquad (2)$$

Alkyl fluorides can be prepared by the addition of hydrogen fluoride to an olefin, Eq. (3). However, the monofluorides are markedly unstable. This is in direct contrast to the polyfluorides in which more than one fluorine atom is attached to the same carbon atom. These are not only stable, but chemically inert as well. It should be noted that monofluoro compounds are frequently quite toxic.

$$CH_2CH_2 + HF \rightarrow CH_3CH_2F \qquad (3)$$

Alkyl halides (iodides, bromides, chlorides) may be made conveniently on a laboratory scale

from the corresponding alcohol by the action of hydrogen halides (anhydrous or concentrated aqueous solution) or of inorganic acid halides, particularly the phosphorus halides. The halogen atom replaces the hydroxyl group of the alcohol, and water is formed as a by-product, Eq. (4). Anhydrous hydrogen halide may be passed directly into the alcohol, or the alcohol may be heated with a concentrated aqueous solution of the acid, usually *in the presence of a catalyst.* The relative rates of reaction of the hydrogen halides are $HI > HBr > HCl$; and of the alcohols, tertiary > secondary > primary alcohol. *Anhydrous zinc chloride* increases the rate of reaction of hydrochloric acid with primary and secondary alcohols. Sulfuric acid catalyzes the reaction of hydrobromic acid with alcohols.

Other halogenating agents include thionyl chloride and the phosphorus halides. *The latter are* used especially for preparing bromides and iodides, particularly of the higher alkanes, from the *respective alcohols*, Eq. (5).

$$ROH + HBr \rightarrow RBr + H_2O \qquad (4)$$
$$3ROH + PBr_3 \rightarrow 3RBr + P(OH)_3 \qquad (5)$$

Alkyl iodides can be produced by the reactions of metallic iodides (sodium or potassium) with other *alkyl halides* (bromides or chlorides) in a *solvent* such as acetone, Eq. (6). The metallic iodides exchange iodine for bromine or chlorine *where the halogen to be exchanged is relatively reactive. Other excellent* methods for the preparation *of alkyl iodides* are *by* the reaction of phosphoric acid, a *metallic iodide,* and an alcohol; and by the gradual addition of red phosphorus to free iodine *in the appropriate* alcohol. The preparation of the *lower-molecular-weight* iodides by this method is hazardous.

$$RBr + NaI \xrightarrow{Acetone} RI + NaBr \qquad (6)$$

of

M

antimony trifluoride is used as the fluorinating agent.

Unsaturated halides. Halogenated *unsaturated* hydrocarbons are classified in two groups, the vinyl *halides and the allyl* halides. The vinyl-type compounds are those in which the halogen *atom is attached* to an unsaturated carbon atom. A typical *example,* vinyl chloride ($CH_2{=}CHCl$), is the *most important of the* unsaturated halogen compounds. It is prepared by *the reaction* of hydrogen chloride with acetylene, using a catalyst, Eq. (7); by *dehydrohalogenation of ethylene* dichloride, Eq. (8); and in the high-*temperature chlorination* of ethane, Eq. (9).

Vinyl chloride is widely used as a monomer in the *manufacture of* polymeric substances, such as plastics and synt fibers. Methods of preparati r to *inyl chloride* are used f l-ty The vinyl compounds are noted for the stability of the carbon halogen bond. *See* VINYL RESIN.

$$HC{\equiv}CH + HCl \xrightarrow{Catalyst} CH_2{=}CHCl \qquad (7)$$

$$ClCH_2CH_2Cl + NaOH \xrightarrow{Heat} CH_2CHCl + NaCl + H_2O \qquad (8)$$

$$H_3CCH_3 + Cl_2 \xrightarrow{High\ temp.} ClCH{=}CH_2 \qquad (9)$$

Allyl halides possess a halogen on a carbon atom adjacent to an unsaturated carbon atom. Typical of the allylics are allyl chloride ($CH_2{=}CHCH_2Cl$) and crotyl chloride ($CH_3CH{=}CHCH_2Cl$). These halides (chlorides, bromides, iodides) are pre*pared from the respective unsaturated* alcohols by the action of the corresponding hydrohalogen acid Eq. (10). Industrially, allyl chloride is produced by the high-temperature chlorination of propylene The iodide may be obtained from glycerol by treatment with either hydriodic acid or a mixture of phosphorus and iodine. The allyl halides are quite reactive and are used extensively as organic intermediates. They are powerful lacrimators.

$$CH_2{=}CHCH_2OH + HCl \rightarrow CH_2{=}CHCH_2Cl \qquad (10)$$

Polyhalogen compounds constitute those halogen-containing hydrocarbons that have two or more atoms of halogen per molecule. These are the polybromo, polychloro, and polyfluoro compounds, together with the mixed halogeno hydrocarbons. The polyiodides are, with few exceptions, difficult to synthesize and generally unstable. The polychlorides and polybromides are produced by the high-temperature chlorination and bromination of the alkanes. Such polychloro compounds as methylene chloride (CH_2Cl_2), chloroform ($CHCl_3$), and carbon tetrachloride (CCl_4) are produced *on a* commercial basis by the controlled *chlorination of* methane. At high temperatures, both substitutive chlorination and combined chlorination and fission of the carbon chain produce *polychlorinated alkanes* and olefins. Reaction products are *generally* carbon tetrachloride, tetrachloroethylene, and hexachloroethane. Ethylene dichloride and propylene dichloride are made from the olefin and gaseous chlorine. Many other polychlorides and *polybromides* may be made by direct halogenation of olefins. This method is used chiefly *for bromination* because the reaction with chlorine is so *vigorous* that it is sometimes explosive. Bromination with liquid bromine goes smoothly at room *temperature* or lower to yield dibromo products with *olefins*, and tetrabromo products with *acetylenes*, Eqs. (11) and (12).

$$RCH{=}CH_2 + Br_2 \xrightarrow{Room\ temp} RCHBrCH_2Br \qquad (11)$$
$$HC{\equiv}CH + 2Br_2 \rightarrow Br_2HCCHBr_2 \qquad (12)$$

Mixed polyhalogen compounds result from halogenation of the unsaturated halides, Eq. (13). Olefinic polyhalogen derivatives are made *by* care-

ful dehydrohalogenation of other polyhalogen compounds by treatment with concentrated sodium or potassium hydroxide solution (aqueous or alcoholic), Eq. (14).

$$CH_2{=}CHCH_2Cl + Br_2 \rightarrow BrCH_2CHBrCH_2Cl \quad (13)$$

$$CHCl_2CHCl_2 + NaOH \xrightarrow{\text{Heat}} CHCl{=}CCl_2 \quad (14)$$

Polyfluorinated hydrocarbons are made by heating chloro or bromo compounds with antimony trifluoride or mercuric fluoride, Eq. (15). The fluorine in the metallic salt exchanges for the halogen atom in the organic compound. The relative ease of replacement follows the general order: I > Br > Cl. Organic polyfluoro compounds, in which the fluorine atoms are attached to the same carbon atom, are more stable than those in which the fluorine atoms are attached to adjacent carbons. Antimony trifluoride is an effective fluorinating agent, particularly if chlorine gas or antimony pentachloride is added to the reaction mixture. About 10% by weight of the antimony trifluoride is recommended. *See* FREON.

$$RCX_2R + SbF_3 \xrightarrow{\text{Heat}} RCF_2R + SbFCl_2 \quad (15)$$

The most important polyiodides are methylene iodide (CH_2I_2) and iodoform (CHI_3). Methylene iodide has the highest density of any liquid except mercury. It is used by mineralogists to determine the density of minerals, and to separate them from each other. It is prepared by the reduction of iodoform with sodium arsenite solution, Eq. (16) Iodoform is produced by the iodination of methyl ketones in aqueous basic solution or by the electrolytic oxidation of ethyl alcohol in aqueous carbonate and iodine solution. Iodoform is used as an antiseptic. *See* HALOFORM REACTION; IODOFORM.

$$CHI_3 + Na_3AsO_3 + NaOH \rightarrow CH_2I_2 + NaI + Na_2AsO_4 \quad (16)$$

Halogenated aromatic hydrocarbons. These include the halogen derivatives of the benzenoid carbon-hydrogen compounds, and are divided into two types, halides in which halogen is attached to a ring carbon, and halides in which halogen is attached to a carbon atom of a side chain The former are called aryl halides; the latter are aralkyl derivatives (benzyl halides), which are similar to the alkyl halides.

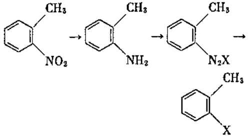
Aryl halides, such as chlorobenzene, and the halogen compounds of naphthalene and anthracene, can be made from the hydrocarbons themselves, or from a derivative of the hydrocarbons such as the nitro or amino. Commercially, benzene is chlorinated or brominated in the presence of a halogen carrier (iron, aluminum, or iodine) at 50°C, Eq. (17). Aluminum chloride, a very active catalyst, produces di- and polyhalogenated products. These can be separated by fractional distillation or crystallization. Chloro derivatives may also be made from benzene and hydrogen chloride by treatment with metallic oxides in air at high temperatures. Iodination with iodine requires the use of a suitable oxidizing agent, such as nitric acid, iodic acid, or persulfate, Eq. (18). If other alkyl groups are present on the ring, reactions must be carried out in the dark to avoid side-chain halogenation. The halogenation of compounds containing a side chain, such as toluene, results in a mixture of the ortho and para isomers, Eq. (19). These can be separated by fractional crystallization. Halogenated aromatic hydrocarbons can be prepared *from amino derivatives of the hydrocarbon (amines* are made by reduction of nitro compounds). This method involves making a diazonium salt from the amine and replacing the salt with halogen using a metallic halide at 0°C, Eq (20). The reaction has the advantage of producing a pure halogen compound, because substitution of halogen occurs only in the position of the amino group. This process is used when there are other alkyl groups present on a ring, and a particular isomer is desired Fluorine, chlorine, bromine, and iodine may be substituted on the ring for an amino group, using the diazonium reaction. *See* DIAZOTIZATION.

$$C_6H_6 + Br_2 \rightarrow C_6H_5Br + HBr \quad (17)$$

$$5C_6H_6 + 2I_2 + HIO_3 \rightarrow 5C_6H_5I + 3H_2O \quad (18)$$

$$CH_3C_6H_5 + Br_2 \xrightarrow{FeBr_3} 1{,}2\text{-}CH_3C_6H_4Br \quad (19a)$$

$$CH_3C_6H_5 + Br_2 \xrightarrow{FeBr_3} 1{,}4\text{-}CH_3C_6H_4Br \quad (19b)$$



(20)

Aralkyl halides, containing halogen in the side chain, *are similar to the alkyl halides. As in the* alkane series, the iodides, bromides, and chlorides can be prepared from the corresponding alcohols by treatment with a suitable halogenating agent

At higher temperatures, it may occur on the second carbon away from the ring (β position), Eq (22). To avoid nuclear substitution, the system must be kept entirely free of iron or iron chloride. Mono-, di-, and trichloro derivatives are obtained. With long side chains, complicated mixtures result, and other methods must be used to obtain the desired product. The phosphorus halides are particularly

useful for preparing the monohalides from alcohols. Fluorine compounds are made by treating halides with antimony trifluoride.

$$C_6H_5CH_2CH_3 + Cl_2 \rightarrow C_6H_5CHClCH_3 + HCl \quad (21)$$
$$C_6H_5CH_2CH_3 + Cl_2 \rightarrow C_6H_5CH_2CH_2Cl + HCl \quad (22)$$

See BROMINE; CHLORINE; FLUORINE; HALOGENATION; IODINE; SUBSTITUTION REACTION.

[E.CL.; M.C.G.]

Halogenation

A chemical reaction in which a halogen atom is introduced into the structure of a molecule or ion in place of some other atom (usually hydrogen) or in place of a group (such as the sulfonic acid group, $-SO_3H$).

Fluorination reactions occur with enormous vigor. In the absence of proper precautions, exposure of organic compounds to elementary fluorine (F_2) results in explosions. Direct reaction of fluorine with organic compounds can be controlled, however, by allowing fluorine and a gaseous organic compound to react in the presence of a copper or silver metal catalyst or by allowing the reaction to occur in a flaming jet. In such reactions, most of the hydrogen atoms of the organic compound are replaced by fluorine atoms and there is often scission of carbon-carbon bonds. Carbon tetrafluoride (CF_4) *is* a common product of fluorination of organic compounds *of any structure* These reactions

by light. However, iodination does not occur. *In alkylbenzenes*, the hydrogen atoms on the side-chain *carbon* immediately next to the nucleus are replaced preferentially. In paraffins, nearly all hydrogens are replaced at significant rates, and a mixture of halogenation products *is obtained*. These reactions also proceed by a radical *mechanism*.

Chlorination, bromination, *and iodination* of aromatic compounds are *electrophilic* substitution reactions. These reactions occur *with varying* ease, depending on the extent to *which the* aromatic compound is activated or *deactivated* by substituents. For halogenation of *highly* deactivated compounds, an effective reagent *is the* molecular

high *electrophilic* re*activity* are formed. For halogenation of *aromatic* compounds of intermediate reactivity, including benzene itself, the molecular *halogen* is used with a catalyst such as $FeBr_3$ or $AlCl_3$; *such* catalysts form complexes (for example, $Br-Br-FeBr_3$) with the halogens, and such complexes are effective in supplying positive halogen *ions* to the aromatic substrates. Strongly *activated* substrates, such as phenols and aromatic *amines*, can be halogenated in water solution in *which* the molecular halogen itself is usually the effective *halogenating agent*.

Because iodination is reversible, it is necessary to prevent the accumulation of hydrogen iodide (HI) in iodination reactions; this can be done by the introduction of some chemical which will react with HI and remove it from the system.

Displacement of sulfonic acid ($-SO_3H$) and carboxyl ($-COOH$) groups by chlorine or bromine occurs readily when these groups are present in an activated position such as ortho or para to a phenolic hydroxy group.

Chlorination, bromination, and iodination of aldehydes, ketones, and esters can be effected at the carbon atom adjacent to the functional group (the α-carbon). Such reactions are catalyzed by either acid or base. If there is no hydrogen atom on the α-carbon atom, as in the case of aromatic aldehydes or of diaryl ketones, this type of halogenation cannot occur. The mechanism of the base-catalyzed reaction involves slow removal of the α-hydrogen by the base to form a carbanion, and then fast reaction of this carbanion with the molecular halogen to form the halogenated product. The mechanism of the acid-catalyzed reaction is similar with the exception that the first step is the slow acid-catalyzed conversion of the ketone or other substrate to an enol of type

$$-\overset{|}{C}=\overset{|}{C}-OH$$

See ADDITION REACTION; HALOGENATED HYDROCARBON; SUBSTITUTION REACTION.

[J.F.B.]

IMPORTANT PRODUCTS

The subject of halogenation and halogenation products is best divided into two groups, the preparation and properties of organic compounds containing fluorine on the one hand and chlorine, bromine, or iodine on the other. The distinction arises from pronounced differences in the preparation of the respective halides and from the unusual general inertness displayed by the organic fluorides toward chemical and thermal attack, whereas the chlorides, bromides, and iodides are considerably reactive. However, not only the chemical properties imparted by fluorine to its compounds are exceptional; fluorine in the molecule tightens the electron cloud, and the low mobility of electrons results in decreased attraction of other molecules. This accounts for the fact that fluoro compounds, when compared not only to compounds containing halogen other than fluorine, but also to the nonfluorinated hydrocarbon analogs, show a far lower boiling point than might be expected by the increase in molecular weight.

Organic fluorine compounds. Selective fluorination is achieved by addition of hydrofluoric acid to a double bond or by exchange of another halogen for fluorine by hydrofluoric acid or a metal fluoride. Substitution of hydrogen by fluorine can be carried out by certain metal fluorides, by hydrofluoric acid which allows perfluorination without attacking the functional group (in ethers, for example), and by elemental fluorine. The latter, however, is a power-

ful agent and renders controlled fluorination difficult. In general, fluorocarbons containing other halogen, or hydrogen, are somewhat less stable. Hydrofluoric acid is the precursor of salts such as potassium hydrofluoride which are used in the industrially important electrochemical fluorination process.

A very conspicuous feature in organic fluorine compounds is the strong inductive effect exerted by the fluorine atoms. This can lead to a decrease in electron density at an adjacent reaction center, and thus influence the chemical reactivity of a given compound. In combination with the aforementioned stability of the carbon-fluorine bond, which renders all parts of the molecule inert except the reaction center (carbonyl, alcohol, and amino group), this influence has helped, and will help, to elucidate the mechanism of many reactions.

Apart from the use as refrigerants of certain short-chain representatives, such as dichlorodifluoromethane, the industrial application of fluoro compounds lay dormant until World War II. The development of atomic power necessitated large-scale production of hydrofluoric acid and intensive investigation of fluoro chemicals and fluorination processes when it was found that only fluorocarbons withstood the corrosive attack of uranium hexafluoride, then used to separate U^{235} from U^{238}. Since then, the rise of fluorine chemistry has been remarkable.

To date, the fluorocarbons most used are starting materials for polymers, such as tetrafluoroethylene (for Teflon) and trifluorochloroethylene (for Kel-F, Fluorothene, Genetron HL). Bifunctional fluorocarbons are the precursors of plastics and other polycondensation products. The polymers and plastics are noteworthy for their general insolubility, their resistance to water, acids, and oxidation, and their high softening points, yet good low-temperature properties. Liquid high-molecular-weight fluorocarbons find uses as lubricants, heat-transfer agents, and tanning agents. Monomeric fluoro compounds are also applicable, for instance in the field of pharmaceuticals; there is an interest in the possibility of their use as anesthetics, and as anticancer agents.

Organic chlorine, bromine, or iodine compounds. Chlorination, bromination, or iodination is generally achieved by substitution of halogen for hydrogen, hydroxyl, or the sulfonic acid group and by addition to unsaturated compounds. As a rule, increasing substitution is accompanied by increasing boiling or melting point, refractive index, and density. Elemental halogens and hydrogen halides are the most frequently used halogenation agents. Numerous other specific reactants have helped to master the art of selective halogenation. Also of specific interest are the chlorinolysis reactions whereby chlorination of hydrocarbons or halohydrocarbons at elevated temperatures results in saturated and unsaturated chlorinated products of molecular weights lower than those of the starting materials. The addition of hydrochloric acid is significant industrially; generally, the organic chlorides are by far the most important halides.

In contrast to organic fluorides, organic chlorides, bromides, and iodides are chemically relatively reactive and therefore are widely used as starting materials in organic synthesis. One of the two general types of reactions most frequently employed makes use of metals or metal compounds (for example, *see* FRIEDEL-CRAFTS REACTION; GRIGNARD REACTION; REFORMATSKY REACTION; WURTZ REACTION), the other is the unimolecular or bimolecular nucleophilic substitution of halogen by other groups. As a generalization, the iodides are more reactive than the bromides, which in turn are more reactive than the chlorides, toward replacement.

In the series of normal saturated halides, the methyl halides are more reactive than the higher homologs. The reactivity of the halogen is markedly enhanced in the allylic halides, where a double bond is adjacent to the carbon carrying the halogen. Vinylic halides, where the carbon carrying the halogen is itself doubly bonded, are remarkably unreactive. Both phenomena are explained by the resonance theory. *See* RESONANCE (MOLECULAR STRUCTURE).

Applications. The industrial application of organic halides is manifold and extensive. Of greatest interest are their uses as starting materials for polymers and plastics, directly as, or as precursors of, insecticides, rodenticides, fungicides, and weed killers, and as solvents or extraction agents. The first two uses apply particularly to chlorohydrocarbons. The polymers and plastics manufactured include Neoprene, from chloroprene; Koroseal and Vinyon, both from vinyl chloride; Saran, a copolymer of vinyl chloride and vinylidene chloride; Dynel, a copolymer of vinyl chloride and acrylonitrile; epon resins, from epichlorohydrin; and ethyl cellulose and silicones, both from ethyl chloride.

Polymers and polycondensation compounds of lower molecular weight serve as detergents, lubricants, flameproofing materials, and fire-extinguishing agents, although (for the latter, monomers such as the bromochloromethanes can be used also. Well-known insecticides, including soil sterilizers and fumigants, are heptachlor, nonachlor, chlordane, Aldrin, Dieldrin, Isodrin, their precursor compounds, dichlorodiphenyltrichloroethane and its related compounds (DDT), the γ-isomer of benzene hexachloride, brominated phenolic compounds, ethylene dibromide, and methyl bromide. Organic halides in their capacity as solvents include carbon tetrachloride (dry cleaning), chloroform, methylene chloride (in paint and varnish removers), methyl chloride (in the production of butyl rubber), tetrachloro- and trichloroethylene (dry cleaning, removal and extraction of greases, fats, and oils). The solvent capacity of organic halides promotes their general use in cleaners of various applicability, polishes, preventives, and sprays. Ethylene dibromide, the most

important bromination product, and ethylene dichloride are added to antiknock fluids to prevent the deposit of lead on the cylinder walls. In the manufacture of tetraethyllead, ethyl chloride is employed. Many organic halides are used for medicinal purposes or serve as intermediates in the synthesis of pharmaceuticals. Thus, chloroform and ethyl chloride are anesthetics; hypnotics are formed from chloral; and diiodohydroxypropane (colorless iodine) is an effective antiseptic. The toxicity and lacrimatory power of organic halides has occasioned their use in chemical warfare. Phosgene, $COCl_2$, and mustard gas, $(C_2H_4Cl)_2S$, are fatal poisons and chloropicrin (nitrochloroform) is a strong tear gas. Specific uses are reserved for chlorofluoromethanes as aerosol propellants, as well as refrigerants, for acetylene tetrabromide as heavy liquid for gages, for iodomethane-benzene mixtures as agents in the separation of minerals, for highly iodinated compounds as x-ray contrast agents, and for allyl chloride in the manufacture of glycerol. See INSECTICIDE; SOLVENT. [E.T.M.]

Hamilton's equations of motion

The motion of a mechanical system may be described by a set of first-order ordinary differential equations known as Hamilton's equations. Because of their remarkably symmetrical form, they are often referred to as the canonical equations of motion of a system. They are equivalent to Lagrange's equations, but the fact that they are of first order and highly symmetrical makes them advantageous for general discussions of the motion of systems. See LAGRANGE'S EQUATIONS.

Definitions. Hamilton's equations can be derived from Lagrange's equations. Let the coordinates of the system be q_j ($j = 1, 2, \ldots, f$), and let the dynamical description of the system be given by the Lagrangian $L(q,\dot{q},t)$, where q denotes all the coordinates and a dot denotes total time derivative (see LAGRANGIAN FUNCTION). Lagrange's equations are then

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = 0 \qquad (1)$$

The momentum p_j canonically conjugate to q_j is defined by

$$p_j = \frac{\partial L}{\partial \dot{q}_j} \qquad (2)$$

It is assumed that these equations (2) are soluble for the velocities $\dot{q}_j$ in terms of the coordinates and momenta. If, as most commonly occurs, the only part of L containing the *velocities* is the kinetic energy, then the canonical momenta are the usual momenta. For example, if $L = m\dot{x}^2/2 - V(x)$, then $p_x = \partial L/\partial \dot{x} = m\dot{x}$.

The Hamiltonian H is defined by

$$H = \sum_{j=1}^{f} p_j\dot{q}_j - L(q,\dot{q},t) \qquad (3)$$

Differentiating Eq. (3) leads to

$$dH = \sum_{j=1}^{f}\left[\left(p_j - \frac{\partial L}{\partial \dot{q}_j}\right)d\dot{q}_j + \dot{q}_j dp_j - \frac{\partial L}{\partial q_j}dq_j\right] - \frac{\partial L}{\partial t}dt$$

$$= \sum_{j=1}^{f}(\dot{q}_j\, dp_j - \dot{p}_j\, dq_j) - \frac{\partial L}{\partial t}dt \qquad (4)$$

where use has been made of Eq. (1) and of Eq. (2) in going from the first to the second line. This shows that H must be a function of the qs, the ps, and t, as the differentials of only these quantities appear in dH. This being so, it follows that

$$dH(q,p,t) = \sum_{j=1}^{f}\left(\frac{\partial H}{\partial q_j}dq_j + \frac{\partial H}{\partial p_j}dp_j\right) + \frac{\partial H}{\partial t}dt \qquad (5)$$

Hamilton's canonical equations result from identifying coefficients of differentials of the independent quantities q_j, p_j, in Eqs. (4) and (5).

$$\dot{q}_j = \frac{\partial H(q,p,t)}{\partial p_j} \qquad \dot{p}_j = -\frac{\partial H(q,p,t)}{\partial q_j} \qquad (6)$$

It is essential that H be written as a function of q, p, t only, and that no velocities appear.

Phase-space coordinates. Hamilton's equations are most easily visualized by introducing the phase space of the system. This is a space of $2f$ dimensions in which the coordinates and momenta of the system serve as the coordinates of a point repre- [illegible] and velocity of each

its position in phase space. An important theorem named for J. Liouville follows directly from Hamilton's equations. Consider several points in a neighborhood of phase space at a given time. These points represent several possible sets of initial conditions for the system. As time proceeds, these points will move. If a large number of points is considered, one may speak of the density of phase points. Liouville's theorem states that the density of phase points in the neighborhood of a given point in phase space does not change as this point moves. The proof is merely the demonstration that the phase points move like an incompressible fluid, that is, the divergence of their velocity vanishes.

$$\sum_{j=1}^{f}\left(\frac{\partial \dot{q}_j}{\partial q_j} + \frac{\partial \dot{p}_j}{\partial p_j}\right) = \sum_{j=1}^{f}\left(\frac{\partial^2 H}{\partial q_j\,\partial p_j} - \frac{\partial^2 H}{\partial p_j\,\partial q_j}\right) = 0 \qquad (7)$$

For additional information on Liouville's theorem and phase space, see STATISTICAL MECHANICS.

Applications. As they stand, Hamilton's equations (6) are no easier to integrate directly than Lagrange's. Elimination of the ps from Eqs. (6) leads back to Lagrange's equations. Hamilton's equations are of great advantage in more general discussions, and they permit the making of canonical transformations which can lead to simplifications. They also lend themselves to numerical integration in some cases where Liouville's theorem

is important, as in ion and electron optics. *See* CANONICAL TRANSFORMATIONS.

The Hamiltonian function H of classical mechanics is used to form the quantum mechanical Hamiltonian operator. *See* QUANTUM THEORY, NONRELATIVISTIC. [P.M.S.]

Bibliography: See LAGRANGE'S EQUATIONS.

Hamilton's principle

A variational statement known as Hamilton's principle forms a basis from which the equations of motion of a classical dynamical system may be deduced. Consider a mechanical system whose configuration is specified by f independent generalized coördinates $q_1, \ldots, q_f$. Let the configuration at time t_1 be given by the values $q_j(t_1)$ and that at time t_2 by the values $q_j(t_2)$. If the system is described dynamically by a Lagrangian function $L(q,\dot{q},t)$ where q stands for all the f coördinates, then, according to Hamilton's principle, the trajectory or path of the system between the two configurations at the two given times is that which makes the value of the integral

$$\Phi = \int_{t_1}^{t_2} L[q(t),\dot{q}(t),t]\,dt$$

stationary relative to nearby paths between the same end points and taking the same time.

Lagrange's equations may be derived from Hamilton's principle as follows.

Let $q_j(t)$ represent the actual trajectory, and let $q_j(t) + \delta q_j(t)$ represent a nearby path. Because the end points are fixed, $\delta q_j(t_1) = \delta q_j(t_2) = 0$. If ϕ is the value of Φ for the trajectory and $\phi + \Delta\phi$ is the value of Φ for the nearby path,

$$\Delta\phi = \int_{t_1}^{t_2} \sum_j \left(\frac{\partial L}{\partial q_j}\,\delta q_j + \frac{\partial L}{\partial \dot{q}_j}\,\delta \dot{q}_j\right) dt$$

Integrating by parts, and noting that

$$\sum_j \frac{\partial L}{\partial \dot{q}_j}\,\delta q_j$$

vanishes at t_1 and t_2,

$$\Delta\phi = \int_{t_1}^{t_2} \sum_j \left(\frac{\partial L}{\partial q_j} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j}\right) \delta q_j\,dt$$

Hamilton's principle asserts that $\Delta\phi$ vanishes for arbitrary δq_j. This can happen only if the coefficient of each δq_j vanishes at all times:

$$\frac{\partial L}{\partial q_j} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} = 0$$

Hence, Lagrange's equations must be satisfied by the coordinates specifying the trajectory of the system.

Hamilton's principle is invariant under any transformation of coordinates. It is also unaffected by the addition of a total time derivative to the Lagrangian, because the integral of such a total derivative is independent of the path of integration. This allows the coordinates q_j and the momenta $p_j = \partial L/\partial \dot{q}_j$ to be subjected to arbitrary canonical transformations without changing the form of Lagrange's equations. *See* CANONICAL TRANSFORMATIONS; LAGRANGE'S EQUATIONS; LAGRANGIAN FUNCTION; MINIMAL PRINCIPLES. [P.M.S.]

Bibliography: See LAGRANGE'S EQUATIONS.

Hamilton-Jacobi theory

A theory that provides a means for discussing the motion of a dynamical system in terms of a single partial differential equation of the first order. It rests on the fact that Hamilton's principle,

$$\Delta \int_{t_1}^{t_2} \left[\sum_{j=1}^{f} (p_j\dot{q}_j) - H\right] dt = 0 \qquad (1)$$

is unaffected by adding a total time derivative to the integrand. If

$$\frac{d}{dt}\,\phi(q,q',t) = \sum_{j=1}^{f} (p_j\dot{q}_j - p'_j\dot{q}'_j) - (H - H') \qquad (2)$$

is subtracted from the integrand, Eq. (1) is reproduced with all quantities replaced by primed quantities. Equation (2) is equivalent to

$$p_j = \frac{\partial \phi}{\partial q_j} \qquad p'_j = -\frac{\partial \phi}{\partial q'_j} \qquad H' = H + \frac{\partial \phi}{\partial t} \qquad (3)$$

See CANONICAL TRANSFORMATIONS; HAMILTON'S PRINCIPLE.

Equations (3) show that it is possible to choose ϕ so that $H' = 0$. Then the equations of motion in the primed variables are simply

$$q'_j = \text{constant} \qquad p'_j = \text{constant} \qquad (4)$$

The description of the motion of the system lies in the function $\phi(q,q',t)$ which generates this transformation. Equations (3) can be solved for q_j, p_j in terms of q'_j, p'_j, and t. The q'_j, p'_j may be chosen as the values of the q_j, p_j at time t_0; then ϕ generates the q_j, p_j at time t from those at time t_0.

The generator of this transformation must be a complete integral of the Hamilton-Jacobi equation

$$H\left(q_j,\frac{\partial \phi}{\partial q_j},t\right) + \frac{\partial \phi}{\partial t} = 0 \qquad (5)$$

which results from Eqs. (3) with $H' = 0$. A complete integral of such a partial differential equation is a solution containing f nonadditive constants of integration which may be taken as the q'_j.

If $H(q,p,t)$ is independent of t so that it may be written $H(q,p)$, then the Hamilton-Jacobi equation is separable with respect to t. Let

$$\phi(q,q',t) = S(q,q') - Et \qquad (6)$$

Then Eq (5) becomes

$$H\left(q_j,\frac{\partial S}{\partial q_j}\right) = E \qquad (7)$$

A complete integral of this equation is required, depending on f nonadditive constants including E.

Schrödinger equation. Equation (5) or (7) bears a close formal resemblance to the Schrödinger wave equation. This resemblance is, in fact

much more than formal. The Hamilton-Jacobi equation is the equation determining the canonical transformation from a set of initial coordinates and momenta to those at time t.

In quantum mechanics, the state vector at time t is found from that at time 0 by a unitary transformation.

$$\psi(t) = U(t)\psi(0)$$

If the coordinates are taken as the quantum numbers describing the state vector $\psi(t)$, then U satisfies the Schrodinger equation.

$$H\left(q,\frac{ih}{2\pi}\frac{\partial}{\partial q},t\right)U - \frac{ih}{2\pi}\frac{\partial U}{\partial t} = 0$$

See QUANTUM THEORY, NONRELATIVISTIC.

Separation of variables. The integration of a partial differential equation is usually more difficult than the solution of a set of ordinary differential equations. It frequently happens, however, that the Hamilton-Jacobi equation is separable in a suitably chosen coordinate system, in which case it may be soluble in practice The general solution is not required, but only a complete integral. Even this is not explicitly required, as only derivatives of ϕ appear in the equation and in the transformation equations which are to be obtained ultimately In separable multiply periodic systems, the frequencies of the motion (as distinct from the trajectory) can be obtained without constructing the transformation function for all times, but just for the periodic times.

The complete integral of the Hamilton-Jacobi equation required is the time integral of the Lagrangian of the system along the trajectory of the system passing through the point q. For, denoting by primes quantities in the transformed coordinate system in phase space,

$$\begin{aligned} L' &= \sum_{j=1}^{f} p'_j\dot{q}'_j - H' = 0 \\ &= L - \frac{d\phi}{dt} \end{aligned} \tag{8}$$

whence

$$\begin{aligned} \phi &= \int_{t_0}^{t} L\,dt \\ &= \int_{t_0}^{t}\left(\sum_{j=1}^{f} p_j\dot{q}_j - H\right)dt \end{aligned} \tag{9}$$

If H is independent of time, so that $H = E$,

$$\phi = \int_{t_0}^{t}\sum_{j=1}^{f} p_j\,dq_j - E(t - t_0) \tag{10}$$

and

$$S = \sum_{j=1}^{f}\int_{t_0}^{t} p_j\,dq_j \tag{11}$$

Equation (10) is especially useful if the system is further separable in the sense that each p_j can be expressed as a function of the conjugate q_j only. Then the problem has been reduced to the evaluation of the integrals occurring in Eq. (10) with arbitrary upper limits (indefinite integrals).

Action variables. If the system is separable and multiply periodic, by which is meant that either each p_k, q_k is a periodic function of the time or that p_k is a periodic function of q_k, it is useful to introduce the action variables J_k. These are defined by

$$J_k = \oint p_k\,dq_k \tag{12}$$

where the integration is over one period of p_k and q_k or is over one period of p_k as a function of q_k. In a separable system, the energy is expressible as a function of the action variables.

$$E = H(J_1, \ldots, J_f) \tag{13}$$

Consider now that Eq. (10) is integrated over a period τ_k of the variable p_k. The function ϕ generates the contact transformation from the initial variables to the variables τ_k later. For the kth degree of freedom, this is

$$\begin{aligned} &p_k \to p_k \qquad q_k \to q_k \\ \text{or}\quad &p_k \to p_k \qquad q_k \to q_k + \text{constant} \end{aligned} \tag{14}$$

In either case, $p_k dq_k$ is the same before and after the transformation, and hence ϕ, according to Eq (2), is independent of the initial and final q_k, and thus contains no reference to the kth degree of freedom; in particular, it is independent of J_k

Equation (10) becomes

$$\phi = J_k + \sum_{\substack{j=1\\ j\neq k}}^{f}\int_{t_1}^{t_1+\tau_k} p_j\,dq_j - H(J)\tau_k \tag{15}$$

The integrals in Eq. (15) are independent of J_k, the term $j = k$ being explicitly missing. Hence, differentiation of Eq. (15) with respect to J_k leads to

$$\begin{aligned} 0 &= 1 - \frac{\partial H}{\partial J_k}\tau_k \\ \nu_k &= \frac{1}{\tau_k} = \frac{\partial H}{\partial J_k} \end{aligned} \tag{16}$$

To calculate the frequencies ν_k of a separable multiply periodic system, it is sufficient to find the energy as a function of the action variables. Equation (16) then gives the frequencies. [P M S]

Bibliography: See LAGRANGE'S EQUATIONS.

Hamster

A rodent, *Cricetus cricetus*, belonging to the family Cricetidae, and a native of temperate Europe and Asia, but most common in Syria. The hamster is about 6 in. long, usually reddish gray above and white below, but it may also be black. It is rather stocky in build, with short ears and a very short tail, being somewhat like the meadow mice in general appearance. In its native habitat it lives in burrows.

Hamsters are favorite animals for laboratory experiments, especially those dealing with nutri-

The hamster, Cricetus cricetus; length to 6 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

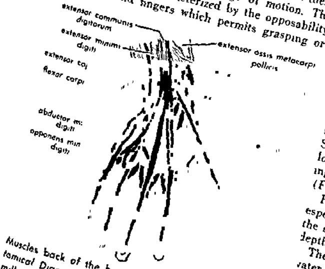
tion, and are reared in captivity in large numbers. Because they are gentle, easily kept, and remarkably free of odor, they also make excellent pets. Hamsters diligently store quantities of grain for winter food, and gather the food by means of cheek pouches. They remain active at temperatures above 45°F; below this temperature they hibernate. See RODENTIA.

[J.D.B.]

Hand

The terminal portion of the upper extremity in primates. The bony framework consists of 8 wrist bones arranged in two rows; 5 palm bones, the metacarpals; and 14 finger bones, the phalanges. These are held together, yet allowed considerable degree of movement, by a complex system of ligaments.

Muscles of the forearm and hand act upon these bones, permitting a wide range of motion. The primate hand is characterized by the opposability of thumb and fingers which permits grasping or



Muscles back of the hand. (From J. M. Dunlop, Anatomical Diagrams for the Use of Art Students, Macmillan, 1935)

prehension. In addition, fine voluntary control affords highly coordinated muscular action in man. The hand also serves as a primary sensory apparatus, because it contains many nervous receptor organs, mediating sensations of touch, pressure, and temperature.

The finer markings of the palmar and finger skin form a sure basis for identification of an individual. See FINGERPRINT.

[E.C.ST.]

Haplosclerida

An order of sponges of the class Demospongiae, including species with a skeleton made up of a single category of siliceous megascleres embedded

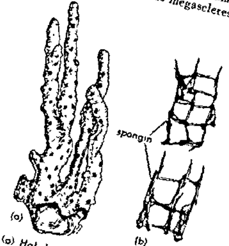


Fig. 1. (a) Haliclona, a haplosclerridan sponge (after Hyman, 1940) (b) Portions of the skeleton of two specimens of Haliclona oculata showing varying amounts of spongin joining the spicules (after Hartman, 1958).

in spongin fibers or joined together in a network by a spongin cement. The megascleres are usually diactinal, sometimes monactinal. Microscleres (never asters) are present or absent. A modified dermal skeleton is absent. Several genera lack spicules and have a skeleton of spongin fibers only. Species of this order are encrusting, massive, or lobate, and many species form large upright branching colonies, the branches often being hollow (Fig. 1).

Haplosclerid sponges inhabit all seas and are especially abundant in tidal and shallow waters of the continental shelf. Some species occur down to depths of 2000 meters.

The family Spongillidae is restricted to fresh water except for a few species which have secondarily invaded brackish water. Spongillids are encrusting, massive, or branching in form and typically inhabit the edges of streams, ponds, and lakes where light shade is present. They are chiefly gray, brown, or white in color. One species is bright yellow and many are green from the occurrence of zoochlorellae in the cells (Fig. 2).

Fossil sponges with a spiculation comparable to that of existing haplosclerids occur scattered through the fossil record from the Cambrian Period on. Fossil spongillids are known from the Jurassic

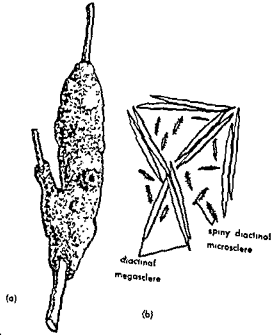
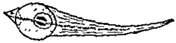


Fig. 2. (a) Spongillid growing on a twig. (b) Spiculation of *Spongilla lacustris*. (After Hyman, 1940)

Period. Spongillid spicules commonly occur in Pleistocene lake sediments. *See* CERACTINOMORPHA; DEMOSPONGIAE. [W.D.H.]

Haplosporidia

A rather ill-defined group of protozoa usually regarded as an order of the Sporozoa. The life cycles of the majority of species are incompletely known. M. Caullery, who with F. Mesnil proposed the order in 1899, now recognizes only 5 genera, which include fewer than 20 valid species. These are parasitic in free-living flatworms, certain flukes (themselves parasitic in bivalve mollusks), mollusks, annelids, crustaceans, and bees. Numerous species of protozoan parasites, thought to be Haplosporidia, have also been described from many other animal hosts. Among these, the hosts are rotifers, insects, primitive chordates, fish, and even man. Certain species have been claimed to be the cause of serious diseases in fish. A more thorough study

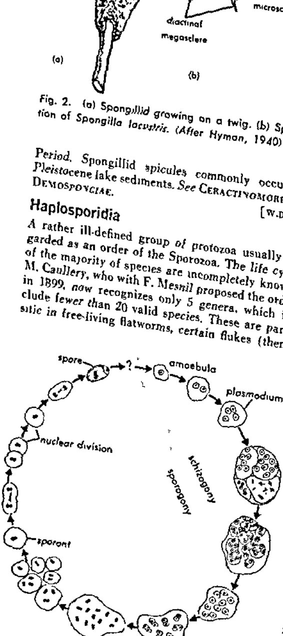


Spore of *Urosporidium fuliginosum*. (After Caullery and Mesnil)

has shown many of these supposed Haplosporidia to be Sporozoa of other kinds, or fungi.

The haplosporidian life cycle begins with a uninucleate spore, lacking polar capsules and polar filaments. These are often of a bizarre shape and represent the infective stage. The spore develops into a multinucleate plasmodium in the coelom or tissues of the host. The multinucleate plasmodium gives rise to uninucleate forms known as sporoblasts, which in turn produce the spores. [R.D.M.]

Bibliography: P. P. Grassé (ed.), *Traité de zoologie*, vol. 1, pt. 2, 1953.



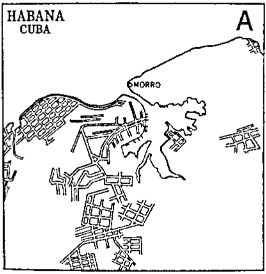
The development of *Ichthyosporidium giganteum* (Swarczewsky).

Harbor

Any body of water of sufficient depth for ships to enter and find shelter from storms or other natural phenomena. The modern harbor is a place where ships are built, launched, and repaired, as well as a terminal for incoming and outgoing ships. There are four principal classes of harbors: commercial, naval, fishery, and harbors of refuge for small craft. Most harbors are situated at the mouth of a river or at some point where it is easy to transfer cargoes inland by river barges, railroads, or trucks.

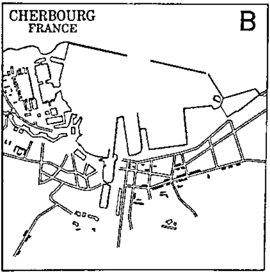
The term harbor is often confused with the term port. A port is a harbor with the necessary terminal facilities to expedite the moving of cargo and passengers at any stage of a journey. A good harbor must have a safe anchorage and a direct channel to open water, and must be deep enough for large ships. An efficient port must have enough room for docks, warehouses, and loading and unloading machinery. Geographically, a port or harbor is usually limited to a comparatively small area of usable berthing space rather than an extended coastline. Some ports along exposed coastal areas, for example, the western coast of South America have very little harbor area.

Types of harbors. There are three types of harbors: landlocked, natural harbors protected from the sea by a narrow inlet; unprotected harbors at which ships may dock even though unprotected from the hazards of changing tides, ocean waves, fogs, and ice; and artificial harbors carved out at sites where the natural features are unfavorable. The latter are fashioned by dredging and by constructing jetties, breakwaters, and sea basins to protect ships against unusually high or low tides.



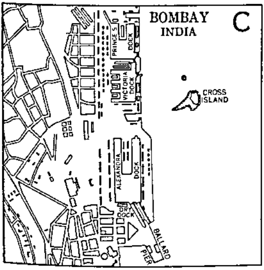
COASTAL (NATURAL)

A coastal harbor sheltered from the wind and sea by virtue of its location within a natural coastal indentation or in the protective lee of an island, cape, reef or other natural barrier



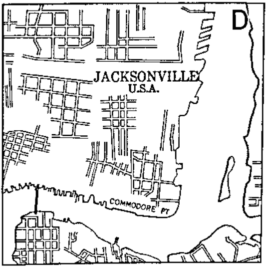
COASTAL (BREAKWATER)

A coastal harbor lying behind a man-made breakwater constructed to provide shelter, or supplement inadequate shelter already provided by natural sources



COASTAL (TIDE GATES)

A coastal harbor, the waters of which are constrained by locks or other mechanical devices in order to provide sufficient water to float vessels at all stages of the tide

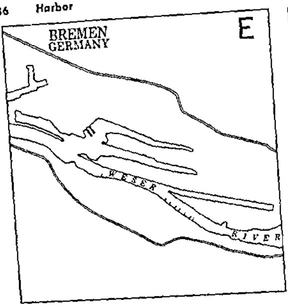


RIVER (NATURAL)

A harbor located on a river, the waters of which are not retained by any artificial means; facilities may consist of quays or wharves parallel to the banks of the stream, or piers or jetties which extend into the stream

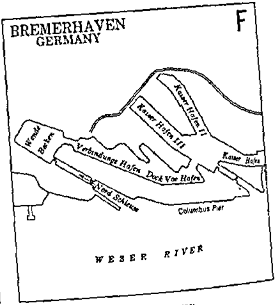
(A–D) Examples of harbor types. (*U.S. Navy Hydrographic Office*)

With the exception of some artificial harbors constructed during World War II, there are few harbors which are entirely man-made. Most harbors are ports, geographically situated at places possessing certain natural and commercial advantages. At these locations, port facilities such as warehouses, piers, quays, docks, and loading and unloading equipment have been constructed and channels have been developed and maintained by constant dredging activity. Some ports supply only fuel and water. The ideal port is deep enough to accommodate the largest cargo freighters and tankers in use. It must have enough room for large ships to maneuver around each other. The bed of



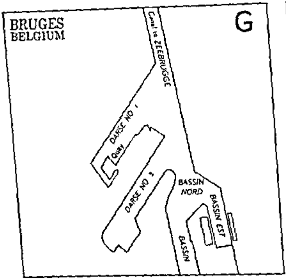
RIVER (BASINS)

A river harbor in which slips for vessels have been excavated in the banks, obliquely or at right angles to the axis of the stream



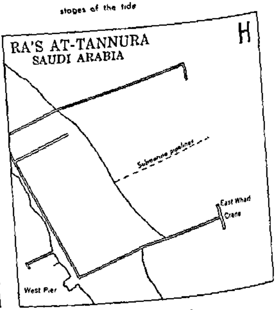
RIVER (TIDE GATES)

A river harbor, the waters of which are constrained by locks or other mechanical devices in order to provide sufficient water to float vessels at all stages of the tide



CANAL OR LAKE

A harbor located in the interior portion of a canal or lake that is connected with the sea by a navigable waterway



OPEN ROADSTEAD

A port which has no natural or artificial barrier to provide shelter from the wind, sea, and swell

(E–H) Examples of harbor types. (U.S. *Navy Hydrographic Office*)

the harbor should not be so rocky, sandy, or muddy that anchors cannot be safely dropped. In latitudes where sea ice is present, harbors should be situated so that shipping operations are curtailed as little as possible during the ice season.

Port construction. Among the most important structures at any modern port are the breakwaters. These structures are barriers designed to intercept or deflect the force of moving water. Their main function is to protect the harbor area from severe wave action. Thus breakwaters provide an artificial means of forming suitable harbors. They also are used to deflect currents and to control silting of channels and deposition of sand on beaches. They must be strong enough to withstand the tremendous impact of waves and should be located so as to avoid detrimental current or wave deflections. Fine examples of ports so protected by break-

waters are Dover, England, and Madras, India. *See* COASTAL ENGINEERING; *see also* SHORE PROCESSES.

Dredging operations are carried out at many ports, especially those located at the mouths of rivers, to keep the channels free of silt deposition and other obstructions which hamper navigation. *See* RIVER ENGINEERING.

Port efficiency. Time is a vital element in computing the effectiveness of a busy port. Loading and unloading, an expensive operation for shippers, has been automated so that ships may quickly discharge their cargoes and take on new loads. Electric cranes, automatic hoists, fast belt loaders, pneumatic tubes (for grains), and rapid pumping facilities (for oil, water, and other liquids) have been developed to speed loading and unloading. Electronic control equipment regulates these operations where previously longshoremen accomplished the same tasks at considerable cost and loss of valuable time. From an economic standpoint a ship resting at a pier is not earning money; consequently, the port which offers the best facilities and makes possible the fastest operation will usually get the traffic. *See* MATERIALS HANDLING MACHINES.

Safety, too, is a prime consideration. Safety in port navigation is as important as safety at the pier. Lighthouses, lightships, buoys, beacons, radiodirectional finders, radar, and sonar are employed collectively along with competent harbor pilots to facilitate safe traffic in a busy port. *See* NAVIGATION SYSTEMS, ELECTRONIC; PILOTING; *see also* BUOY; LIGHTHOUSE.

Aside from its wharves, piers, and cargo-handling machinery, the modern port must also possess rail and truck terminals to facilitate the transfer of cargo from one mode of transportation to another. This includes sufficient warehousing facilities to handle the storage of cargo awaiting shipment. Other aspects and requirements of modern ports are police protection to prevent pilferage at the piers, auxiliary vessels such as tugs and water taxis, and facilities to handle passenger traffic. Provision must also be made for customs inspections and banking and other commercial enterprises which handle the paper work involved in marine insurance, finance, and the preparation of shipping documents. [N.W.]

Hardness scales

Arbitrarily *defined measures of the resistance of a* material to *indentation under static or* dynamic load, to scratch, *abrasion, or wear,* or to cutting or drilling. Standardized tests compare similar materials according to the particular aspect of hardness measured by the test. Widely used tests for metals are Brinell, Rockwell, and Scleroscope tests, with modifications depending upon the size or condition of the material. Indentation tests compare species of wood or flooring materials, and abrasion tests serve as an index of performance of stones and paving *materials.*

Hardness tests are important in research and are widely used for grading, acceptance, and qu control of manufactured articles. The hardness designation or scale is associated with the test method or instrument used.

Scratch hardness. Resistance to scratching is defined by comparison with ten selected minerals, called Mohs scale, which are numbered in the order of increasing hardness. This mineralogical scale is 1, talc; 2, gypsum; 3, calcite; 4, fluorite; 5, apatite; 6, orthoclase; 7, quartz; 8, topaz; 9, corundum; and 10, diamond. Minerals lower in the scale are scratched by those with higher numbers. The scale is extended to provide finer distinction of harder materials by additional minerals: 7, vitreous pure silica; 8, quartz; 9, topaz; 10, garnet; 11, fused zirconia; 12, fused alumina; 13, silicon carbide; 14, boron carbide; and 15, diamond.

File-test hardness. Materials are differentiated qualitatively according to resistance to scratching or cutting by files especially selected for the purpose. Whether or not a visible scratch is produced on the material indicates its hardness in comparison with a sample of desired hardness. The method is used for routine inspection of hardened surfaces in production.

500 kg or less for soft materials. Various machines apply and control the specified load. The diameter of the impression is measured with a micrometer microscope Brinell hardness number (Bhn), expressed in kg/mm², is obtained by dividing the load by the spherical surface area of the impression Different size balls may be used according to

Applied load may be varied from 5 to 120 kg in increments of 5 kg according to size of test piece. Vickers hardness number, also called diamond pyramid hardness, is equal to the load divided by the lateral area of the pyramidal impression. The area is computed from measurements of the diagonals of the square impression. Vickers hardness is the most reliable measure for very hard material and is applicable to thin sheets and hardened surfaces.

Rockwell hardness. Depth of indentation of either a steel ball or a 120° conical diamond with rounded point, called a brale, under prescribed load is the basis for Rockwell hardness. The ball is normally 1/16 in. in diameter, but 1/8, 1/4, or 1/2-in. balls are used for soft materials. A special signed machine applies loads of 60, 100,

The depth of impression, referred to the position under an initial minor load, is indicated on a dial whose graduations represent the hardness number. Hardness is designated by a number with a standard system of prefix letters to indicate type of penetrator and load used.

Superficial Rockwell hardness is measured by a special machine differing from the standard Rockwell tester in that it applies lighter loads with a more sensitive depth-measuring dial. It produces a shallow impression and is suitable for thin sheet material and where surface hardness to a limited depth is of interest.

Monotron hardness. The pressure in kg/mm² required to embed a 0.75-mm hemispherical diamond penetrator to a depth of .0018 in., producing an impression 0.36 mm in diameter, is the measure of Monotron hardness. The depth is controlled by a separate dial graduated to 1 kg/unit area. The method is applicable to the entire range of hardness and is suited to thin sheet and casehardened surfaces.

Shore Scleroscope hardness. Height of rebound of a diamond-tipped weight or hammer falling within a glass tube from a height of 10 inches and striking the specimen surface measures Shore Scleroscope hardness. The standard hammer is ¼ in. in diameter, ¾ in. long, and weighs 1⁄12 oz. The hardness number is the height of rebound referred to an arbitrary scale graduated to 140 divisions within the glass tube. The method is a dynamic load test, and the rebound reflects the size of indentation produced, which determines the energy absorbed by deformation and hence that available for rebound. A dial-recording instrument indicates the rebound hardness directly. Both instruments are portable and permit rapid determinations.

Herbert pendulum hardness. Resistance to cold working is measured as Herbert pendulum hardness. The apparatus consists of a rocking device, called a pendulum, supported on a 1-mm steel ball in contact with the specimen. A curved level bubble measures amplitude of oscillation. The time hardness number is the time in seconds for 10 complete swings of the pendulum through a small arc. The work-hardening capacity is measured by the maximum time hardness after previously repeated single swings of the pendulum. The scale hardness number is the angular oscillation of a half swing after tilting the pendulum through a definite angle before release. Scale hardness is taken as a measure of resistance to flow as in rolling, drawing, and stamping. The method is applicable to studies of machining and forming of metals.

Microhardness. Resistance to indentation over very small areas (as on small parts, the constituents of metal alloys, or for exploration of hardness variations) is called microhardness. One tester employs the Vickers square-based pyramidal diamond penetrator attached to the end of a vertically guided shaft having a weight of 25 g. An arrangement of microscopes permits centering and measurement of the diagonals of the impression. The hardness number is the pressure intensity in kg/mm².

Another procedure employs a Tukon tester applying loads of 25–3600 g using a Knoop indenter, which is a diamond ground to produce a diamond-shaped impression with ratio of diagonal lengths of 7 to 1. The location of the indenter and measurement of the diagonals of the impression is accomplished with microscopes. The hardness number is the ratio of the applied load to the projected area of the impression.

Hardness of wood. The load required to embed a 0.444-in. diameter steel ball to half its diameter expresses the hardness of wood. Used as a means of comparison, the values vary with species and grain characteristics. Hardness values of poplar and Douglas-fir are approximately 400 and 900 lb respectively.

Hardness of paving. Wear or abrasion hardness applies primarily to natural stones, paving, or flooring materials. It is measured by a specified test providing an index of service performance. Hardness of stone is reflected by weight loss of a cylindrical core rubbed on a sand bed. Deval abrasion test determines weight loss of a charge of broken stone tumbled in a cylinder. Similarly, the Los Angeles rattler and the standard rattler test for paving block tumble a charge including steel balls in a drum to determine percentage loss of weight as an index of wear. Special wear tests are applied to floor surfaces. [W.J.KR.]

Bibliography: H. E. Davis, G. E. Troxell, and C. T. Wiskocil, *Testing and Inspection of Engineering Materials*, 2d ed., 1955; G. L. Kehl, *Principles of Metallographic Laboratory Practice*, 3d ed., 1949.

Hardy-Weinberg formula

A basic mathematical relation used in population genetics. It gives the proportion of the various genotypes in a randomly mating population in terms of the frequencies of the genes. The formula is useful for genetic analysis of populations, such as human populations or plants and animals in nature ... are not possible.

... g,

a German physician who made a number of important contributions to the methodology of human genetics. *See* HUMAN GENETICS; POPULATION GENETICS.

In its simplest form the Hardy-Weinberg formula may be stated this way: if p is the proportion of gene A in the population and q $(= 1 - p)$ is ... from one generation ... a, and ... nd q^2.

In other words the genotypes are given by the appropriate terms in the expansion of the binomial $(p + q)^2$.

As a numerical example, the recessive gene a for phenylketonuria, which is a metabolic deficiency re-

sulting in feeble-mindedness, has a frequency in the United States of about 0.006. Therefore the proportions of the three genotypes are

Homozygous normal	AA	$(0.994)^2$	$= 0.988036$
Heterozygous normal	Aa	$2(0.994)(0.006)$	$= 0.011928$
Feeble-minded	aa	$(0.006)^2$	$= 0.000036$

Notice that there are about 330 times as many persons who are heterozygous carriers (Aa) of the gene as there are persons homozygous (aa) for it. This is characteristic of rare recessive factors.

The formula holds only for an infinite population and assumes random mating in the absence of significant mutation pressure or gene transfer between populations. However, it is an accurate approximation in many populations. Random mating in this context means that matings occur without regard to the characters determined by the genes in question or the degree of relationship of the mates. This means that it is possible for a population to be mating at random with respect to some genes and not for others at the same time. For example, it is appropriate to regard the human population as mating at random for blood group genes, and the data actually show excellent agreement with Hardy-Weinberg predictions. However, the formula would not be expected to hold for genes that determine such characters as skin color or intelligence, which strongly influence the choice of mates.

Gene frequency analysis is a technique for testing genetic hypotheses in randomly mating populations. The Hardy-Weinberg formula or an extension of it is used to predict the frequency of certain types in the population or in the progeny of certain parental combinations and these are compared with the frequencies actually observed. See GENETICS.

[J.F.C.]

Bibliography: C. C. Li, *Population Genetics*, 1955.

Harelip

A form of congenital defect, sometimes hereditary, marked by abnormal clefts between the upper lip and the base of the nose. Most of these clefts are asymmetrical, occurring on either side of the midline (Fig 1). Occasionally they are bilateral

Many cases of harelip occur in association with other defects of the facial area, most commonly those of cleft palate (Fig. 2).

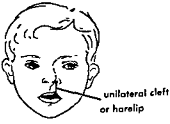


Fig. 1. A harelip. (From L. B. Arey, *Developmental Anatomy*, 6th ed., Saunders, 1954)

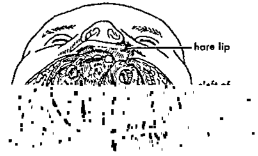


Fig. 2 Cleft involving total soft and hard palates, also combined with severe cleft lip. (From L. B. Arey, *Developmental Anatomy*, 6th ed., Saunders, 1954)

In all of these disorders an incomplete formation of the central structures of the face is the cause. It follows a failure of mesodermal tissue to penetrate and replace the epithelial junction between the paired processes of the maxilla, or upper jaw bone.

Early surgery during the first few months of life is indicated for cleft lip, or harelip. If the defect extends into the palate, the size of the opening and the tissues involved will generally dictate the most opportune time for repair.

In any case, corrective measures should be undertaken before the child begins to talk, since phonation, or speech, is severely affected, particularly in combination defects with penetration through the oral cavity into the nasal chambers. See SPEECH.

[E.C.ST.]

Harmonic (periodic phenomena)

A sinusoidal quantity having a frequency which is an integral multiple of the frequency of a periodic quantity to which it is related. See MODE OF VIBRATION; PARTIAL TONE.

A harmonic series of sounds is one in which the basic frequency of each sound is an integral multiple of some fundamental frequency. The name exists for historical reasons, even though according to the usual mathematical definition such frequencies form an arithmetic series. An ideal string (or air column) can vibrate as a whole or in a number of equal parts, and the respective periods of vibration are proportional to the lengths. These increasingly shorter lengths or periods form a harmonic series. The name came from the harmonious

in acoustics becomes that given here. See HARMONIC ANALYZER; MUSICAL ACOUSTICS. [R.W.Y.]

Harmonic analyzer

A device for separating and measuring the frequencies and amplitudes of the Fourier-series components of a complex periodic wave. See WAVEFORM, NONSINUSOIDAL.

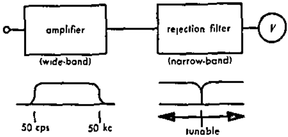


Fig 1 Basic distortion and noise meter.

Wattmeter measurement. At low frequencies, harmonics may be measured with a wattmeter. The voltage to be analyzed is impressed upon the potential coil, and a current of adjustable frequency is passed through the current coil. When the frequency of the current I is set approximately equal to that of a harmonic component E_t of the voltage, the meter will deflect to indicate the product $E_tI \cos \theta$, where θ is the phase angle between the voltage and current. As θ varies through 360° during the period of a beat between the two frequencies, the wattmeter will swing between the maximum positive and negative readings for which $\cos \theta = \pm 1$ By adjusting the frequency of the current to be close to that of the harmonic, a very slow beat can be produced and an accurate maximum reading obtained. The instrument can be calibrated, for any given amplitude and frequency of the search current, by impressing a sine-wave voltage of known amplitude on the potential coil and adjusting it in frequency and amplitude to produce the same maximum deflection. This method is simple and accurate, even for small harmonic components, but is limited to the lower audio frequencies for which dynamometer-type instruments are suitable *See* WATTMETER.

Tunable filters. For wider frequency ranges tunable filters can be used to select or reject signals that fall within narrow bands. Two widely used types of instrument have been developed commercially that make use of such filters. The first of these, generally referred to as a distortion and noise meter, is designed to amplify uniformly over the frequency range of interest. A narrow band rejection filter, which can be tuned to the fundamental frequency of the signal to be analyzed, is used [illegible]

a suitable vacuum-tube voltmeter calibrated in per cent noise, usually expressed in decibels (db). It is widely used in maintaining broadcast station circuits in proper operating condition.

The second type of instrument is shown in Fig 2 A tunable narrow-band rejection filter is again used, but in this application it reduces the negative feedback in a heavily degenerated amplifier to a minimum at its rejection frequency. The amplifier then has maximum gain and will select any signal occurring at this frequency while rejecting others. Instruments of this type are often used for analysis of noise spectra, as well as for harmonic selection and measurement. For a general analysis of this nature an important characteristic is that the passband is a constant fraction of the tuning frequency.

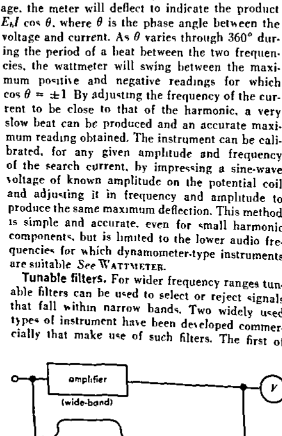


Fig. 2. Sound and vibration analyzer.

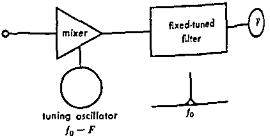


Fig 3. Wave analyzer.

Heterodyne measurements. Another widely used instrument employs a highly selective fixed tuned filter in the heterodyne system illustrated in Fig. 3. To produce a highly selective filter at a frequency well above the highest frequency to be measured, quartz crystals are used as the frequency-selective elements. The heterodyning oscillator covers a frequency range from the center frequency of the filter f_0 to a frequency $f_0 - F$, making possible measurements at any frequency from 0 to F.

Instruments of this type, generally called wave analyzers, are used for a wide variety of measurements requiring high precision and accuracy. In contrast to instruments using tunable filters, they maintain a measurement bandwidth Δf that is constant, independent of frequency. They are therefore particularly useful for measuring noise spectra where the energy content per cycle of bandwidth is important. *See* ELECTRICAL MEASUREMENTS; FREQUENCY MEASUREMENT. [illegible]

Bibliography: H. H. Scott, A new type of selective circuit and some applications, *Proc. IRE*, 26(2):226–235, 1938; F. E. Terman and J. M Pettit, *Electronic Measurements*, 2d ed , 1952.

Harmonic motion

A periodic motion that is a sinusoidal function of time. It is often called simple harmonic motion (SHM). It is the simplest possible type of vibratory motion. The motion is symmetric about its midpoint, at which the velocity is greatest and the acceleration is zero. At the extreme displacements or turning points, the velocity is zero, and the acceleration is a maximum. The motion is characterized by a unique frequency (without overtones).

Harmonic motion may be present in very simple mechanisms. For example, if a wheel is rotating at constant speed about a fixed axis, the projection on any fixed line of the motion of a point on the wheel is simple harmonic. Harmonic motion may also result from the response of a vibrating system to a periodic—in particular a sinusoidal—force. Harmonic motion is the typical motion of most simple systems that have been displaced from a position of stable equilibrium and then released, provided that the damping is negligible. The motion of a pendulum is approximately simple harmonic for small amplitudes. *See* PENDULUM.

If x represents the displacement measured from the midposition, and t the time, then harmonic motion can be described by either of the forms

$$x = A\cos(\omega t) + B\sin(\omega t)$$
or
$$x = C\sin(\omega t - \delta) \qquad (1)$$

The constants A, B, C, and δ are not all independent, but have the following relations:

$$A = -C\sin\delta \qquad B = C\cos\delta$$

The amplitude C represents maximum displacement in one direction from the center (half the total motion between extreme positions), δ is a phase angle whose value depends on the precise instant at which the oscillation was started, or alternatively on the phase of the motion when $t = 0$. These quantities are illustrated in Fig. 1.

The remaining constant, ω, is known as the angular frequency. Dimensionally, ω is the reciprocal of time. Thus the product ωt is a pure number, to be interpreted as an angle measured in radians. When ωt increases by 2π, the motion repeats. Thus the angular frequency ω is related to ordinary frequency f (number of complete oscillations in unit time) and period T (duration of one complete oscillation) by Eq. (2).

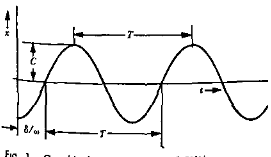


Fig. 1. Graphical representation of SHM.

$$\omega = 2\pi f = 2\pi/T \qquad (2)$$

The velocity v and acceleration a, obtained by differentiating Eq. (1), are given by

$$v = dx/dt = \omega C\cos(\omega t - \delta) = \omega\sqrt{C^2 - x^2} \qquad (3)$$
$$a = dv/dt = -\omega^2 C\sin(\omega t - \delta) = -\omega^2 x \qquad (4)$$

Because the net force acting on a body is equal to the mass of the body multiplied by its acceleration, Eq. (4) shows that in SHM, the force must be proportional to the displacement.

$$F = ma = -m\omega^2 x \qquad (5)$$

Conversely, SHM occurs whenever, for a body that is displaced from an equilibrium position, there is a net restoring force proportional to the displacement.

Energy considerations. The importance of harmonic motion lies in the simplicity of its time dependence, as given by Eq. (1), and in the frequent occurrence of linear restoring forces. For sufficiently small displacements of almost any mechanical system from equilibrium, the restoring force or torque is always approximately proportional to the displacement.

This can be explained by some considerations about potential energy. At a point of stable equilibrium, potential energy is necessarily a minimum. If potential energy is a well-behaved function of position, it may be expanded as a series of powers of *the distance from any point. (An important* exception is electrostatic potential, which varies inversely as the distance from the charge, and so is not well-behaved in the immediate neighborhood of the charge.) In the expansion of potential energy about a point of stable equilibrium, the special minimum property guarantees that the linear term must vanish. The next term, quadratic in the distance from the equilibrium point, does not ordinarily vanish, and provides a good approximation of the potential energy for sufficiently small displacements. A quadratic, or parabolic, variation of potential with distance is equivalent to a force that varies linearly.

In a system oscillating freely about an equilibrium position, the energy changes from potential to kinetic and back again. For SHM, the potential energy, proportional to the square of the displacement, is greatest at the extremities of the motion. Conversely, the kinetic energy $\frac{1}{2}mv^2$ is zero at the

motion, and the average potential energy just equals the average kinetic energy.

The frequency of a freely oscillating system is determined by the stiffness and inertia of the system. If the motion is harmonic, it is also isochronous, which means that the frequency is independent of the amplitude of the motion.

Weight on elastic spring. If an elastic spring is a distance x beyond its natural l...

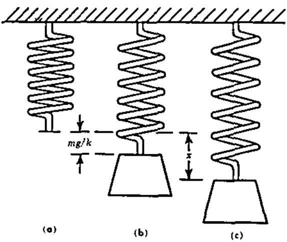


Fig. 2 Weight on an elastic spring. (a) Unloaded spring. (b) Statically loaded spring. (c) Weight and spring oscillating

or compressed a distance x short of its natural length, it exerts a restoring force equal to $-kx$. The stiffness of the spring is measured by the spring constant k. Consider a spring which is suspended vertically and set into vertical oscillations, with a mass m attached to its free end (see Fig. 2). If m is large compared to the mass of the spring, the latter can be neglected.

The equation of motion for the mass m is

$$mg - kx = ma = m\,d^2x/dt^2$$

where g is the acceleration of gravity. The solution of this equation can be written in the form

$$x = mg/k + C \sin(\omega t - \delta)$$

with $\omega^2 = k/m$

The term mg/k is simply the extension of the spring due to the static weight, and marks the loaded equilibrium position. Displacement from this equilibrium position calls forth a linear restoring force, and the resulting motion is simple harmonic.

Potential energy (PE) for this example is the sum of an elastic energy $\frac{1}{2}kx^2$ and a gravitational energy $-mgx$

$$\begin{aligned} \mathrm{PE} &= \tfrac{1}{2}kx^2 - mgx \\ &= \tfrac{1}{2}k(x - mg/k)^2 - m^2g^2/2k \end{aligned}$$

showing the quadratic dependence on displacement from equilibrium. The constant term $-m^2g^2/2k$

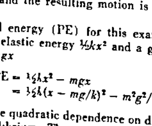


Fig 3. Torsional pendulum.

is of no importance to the motion; it can always be removed by a new choice for the zero of potential energy.

Rotational harmonic motion. Rotational as well as translational SHM may occur. In this case the angular displacement is a sinusoidal function of time. A free angular vibration will be simple harmonic if the angular displacement from an equilibrium orientation produces a restoring torque proportional to the displacement. An example is the torsional pendulum (Fig. 3). One end of a torsionally flexible elastic rod is held fixed. To the other end is fastened a disk, or another body of large moment of inertia. If the rod is twisted and then released, the rod and disk will undergo angular SHM, provided that the torque in the rod is proportional to the angle of twist.

Atomic vibrations. The realization that atoms are continually vibrating in motions that are nearly harmonic is essential for understanding many properties of matter, including molecular spectra, heat capacity, and heat conduction.

In a diatomic molecule, the distance between the atoms is not precisely fixed. There is an equi-

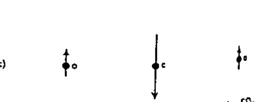


Fig. 4. Illustrating three normal modes of the CO_2 molecule. (a) Lower frequency in-line mode. (b) Higher frequency in-line mode. (c) Bending mode.

librium separation corresponding to a minimum in potential energy. The actual magnitude of the interatomic distance oscillates about the equilibrium distance with a motion that is approximately simple harmonic, if the energy of oscillation is small.

In a polyatomic molecule or in a crystalline solid, the atoms also vibrate about equilibrium positions. (For a description of this phenomenon in crystalline solids, *see* LATTICE VIBRATIONS.) Because of the large number of degrees of freedom, however, the situation is more complicated. For example, in the carbon dioxide (CO_2) molecule neither oxygen atom by itself moves harmonically, or even periodically. On the other hand, the motion of the CO_2 molecule can be analyzed into a number of independent motions, called normal modes, each of which is by itself simple harmonic. In one such motion, for example, the two oxygen atoms move in phase toward or away from the carbon

atom (see Fig. 4a). The actual motion of the atoms in a molecule is a superposition of the various normal modes and a rotation of the molecule as a whole. *See* MOLECULAR STRUCTURE AND SPECTRA. *See also* DAMPING; FORCED OSCILLATION; HARMONIC OSCILLATOR; PERIODIC MOTION; VIBRATION; WAVE MOTION.

[J.M.KE.]

Bibliography: K. R. Symon, *Mechanics*, 1953.

Harmonic oscillator

Any physical system that is bound to a position of stable equilibrium by a restoring force or torque proportional to the linear or angular displacement from this position. If such a body is disturbed from its equilibrium position and released, and if damping can be neglected, the resulting vibration will be simple harmonic motion, with no overtones. The frequency of vibration is the natural frequency of the oscillator, determined by its inertia (mass) and the stiffness of its restoring force.

The harmonic oscillator is not restricted to a mechanical system, but might, for example, be electric. Typical electronic oscillators, however, are only approximately harmonic.

If a harmonic oscillator, instead of vibrating freely, is driven by a periodic force, it will vibrate harmonically with the period of the force; initially the natural frequency will also be present, but any damping will eventually remove the natural motion. The response of a harmonic oscillator driven by a general force $f(t)$, an arbitrary function of time, is described by a linear differential equation. If $x(t)$ represents the displacement as a function of time t, then the equation of motion is

$$m\,d^2x/dt^2 = -c\,dx/dt - kx + f(t) \qquad (1)$$

The left side of Eq. (1) represents the mass m multiplied by the acceleration. The force on the right side of the equation includes, in addition to the restoring force $-kx$ and the driving force, a "viscous damping" force $-c\,dx/dt$ that is proportional to the velocity. (This damping may vanish.) The explicit solution of Eq. (1) is

$$x = e^{-bt}[A\sin(\omega t - \delta) + \frac{1}{m\omega}\int_{-\infty}^{t} e^{bT}\sin[\omega(t-T)]f(T)\,dT] \qquad (2)$$

where $b = c/2m$, $\omega^2 = (k/m) - (c/2m)^2$, and A and δ are arbitrary constants which can be set equal to zero if the oscillator displacement and velocity were zero before application of the force $f(t)$. The variable of integration T is the time at which the force $f(T)$ is considered to act. *See* DAMPING; FORCED OSCILLATION; HARMONIC MOTION.

In both quantum mechanics and classical mechanics, the harmonic oscillator is an important problem. It is one of the few rigorously soluble problems of quantum mechanics. The quantum mechanical description of electromagnetic, electronic, mesonic, and other fields is usually carried out in terms of a (time) Fourier analysis. The individual Fourier components of noninteracting fields are independent harmonic oscillators.

The Hamiltonian for a harmonic oscillator is

$$H = p^2/2m + m\omega^2q^2/2 \qquad (3a)$$
$$= P^2/2 + \tfrac{1}{2}\omega^2Q^2 \qquad (3b)$$

where ω is the angular frequency characteristic of the oscillator. The mass m in Eq. (3a) is made to disappear by replacing the coordinate q and the conjugate momentum p by

$$Q = m^{1/2}q$$
$$P = m^{-1/2}p$$

in Eq. (3b). (For explanation of the term conjugate momentum, *see* LAGRANGE'S EQUATIONS.) The corresponding Schrodinger equation for the wave function ψ is then

$$-(\hbar^2/2)\,d^2\psi/dQ^2 + (\omega^2Q^2/2)\psi = E\psi \qquad (4)$$

where E is the energy, and $\hbar$ is Planck's constant divided by 2π. This equation possesses quadratically integrable solutions only for the characteristic energy values

$$E = (n + \tfrac{1}{2})\hbar\omega \qquad (5)$$

where n is any positive integer. For these cases, ψ can be expressed in terms of the Hermite polynomial $H_n(y)$ of degree n, as

$$\psi_n = C_n e^{(-y^2/2)}H_n(y) \qquad (6)$$

where $y = (\omega/\hbar)^{1/2}Q$

is dimensionless, and

$$C_n^2 = \left(\frac{\omega}{\pi\hbar}\right)^{1/2}\frac{1}{2^n n!}$$

See ANHARMONIC OSCILLATOR; QUANTUM THEORY, NONRELATIVISTIC.

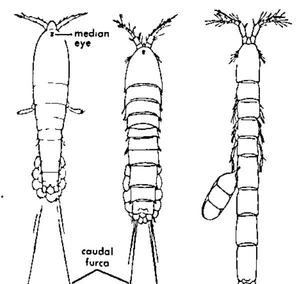
[J.M.KE.]

Bibliography: J. P. Den Hartog, *Mechanical Vibrations*, 4th ed., 1956; J. W. S. Rayleigh, *The Theory of Sound*, 2d ed., reprint, 1945; L. I. Schiff, *Quantum Mechanics*, 2d ed., 1956.

Harpacticoida

An order of the Copepoda derived from the genus *Harpacticus* Milne-Edwards 1840. This group of crustaceans is referred to as Podoplea Copepoda. Their form is variable but is generally linear and more or less cylindrical. These animals are minute and vary in length from 0.4 to 3 mm. As a rule, the first thoracic segment is incorporated with the cephalothorax, and the last thoracic segment is included in the abdomen. Antennulae in the males are prehensile, while the antennae in both sexes are usually biramous. The mouth parts of free-living forms are well developed, while in parasitic forms they are reduced or transformed. Maxillipeds as a

almost always three-jointed, and endopodites three- to one-jointed or absent. Sexual dimorphism can occur in all legs mentioned (*see* SEXUAL DIMORPHISM). The last pair of legs is reduced and dis-



(a) *Harpacticus chelifer.* (b) *Ameira longipes.* (c) *Cylindropsyllus laevis.* (d) *Anchorabolus mirabilis* (e) *Parategastes spaericus.* (f) *Porcellidium ravanae.* (g) *Alteutha interrupta.*

similar in the two sexes. The ovisac may be single or double, but is always ventral. Ova of a few species are laid free.

The nervous system consists of a supra- and a subesophageal ganglion, a circumesophageal ring and a [illegible] segme[illegible] the ey[illegible] [illegible] they are present, the frontal organs. The circumesophageal ring innervates the dorsal musculature of the cephalothorax, the antennae, and the labrum, while the subesophageal ganglion innervates the rest of the mouth parts. The last-mentioned ganglion is provided with two transverse commissures.

As a rule a tripartite median eye, the nauplius eye, is present. The parts of the eye, however, may be independent, or the eye may have disappeared. One pair of frontal organs is sometimes to be found. Antennulae have an olfactory organ, the aesthete, on the fourth joint or its equivalent; the males often have aesthetes on other joints also.

The gonads are single. Their ducts open ventromedially on the first true abdominal segment. Two oviducts open into a receptaculum seminis. The vas deferentia are single or paired, and spermatozoa are contained in spermatophores.

The alimentary canal is similar to that in most other copepods. Excretory organs consist of a maxillary gland which is always present, and a rudimentary antennal organ occurs in some forms. Development follows the general plan as it occurs in other copepods. *Elaphoidella bidens* exhibits parthenogenesis, while heterogony is found in *Epactophanes*. Harpacticoids have a world wide distribution. They occur in all kinds of aquatic habitats, especially marine, but also among moss and leaves. In the sea they range from the shore to abyssal depths. In general, they are free-living and benthonic. Some few species are pelagic, parasitic, or commensal. The species may be mono- or polycyclic. A few fresh-water species produce resting eggs, or they estivate within cysts About 1400 species are known. *See* COPEPODA.

[K.L.]

Bibliography: K. Lang, *Monographie der Harpacticiden*, 1948; G. O. Sars, *Crustacea of Norway*, vol. 5, 1911; G. O. Sars, *Crustacea of Norway*, vol. 7, 1921.

Hatchettite

Mountain tallow, a yellow-white to yellow-green hydrocarbon occurring in Belgian coal seams. The material, also called hatchettine, frequently is found in geodes, also in fractures in associated rocks. Hatchettite is translucent but darkens on exposure to air. It is soft, has no odor, is greasy to the touch, and consists of 85.5% carbon and 14.5% hydrogen leading to the empirical formula C_nH_{2n}. Its index of refraction is 1.47–1.50, it melts at 46–47°C, is sparingly soluble in alcohol or ether, decomposes in concentrated sulfuric acid, and has a specific gravity of 0.89–0.98.

[L.A.B.]

Hausmannite

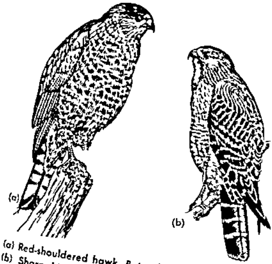
A mineral having composition $MnMn_2O_4$ and crystallizing in the tetragonal system. Hausmannite crystals are pyramidal pseudooctahedral, but the mineral is commonly massive. Cleavage is nearly perfect parallel to the basal pinacoid. The hardness is 5.5 (Mohs scale) and the specific gravity 4.81. The luster is submetallic and the color brownish black. Hausmannite has been reported from many widely scattered localities. It occurs in metamorphosed sedimentary manganese ore, in high-temperature hydrothermal veins, and in contact metamorphic deposits. *See* MANGANESE; MANGANITE.

[C.S.H.]

Hawk

Any of a large number of diurnal birds of prey of the order Falconiformes, an order that includes the hawks, vultures, and eagles (*see* EAGLE; VULTURE).

Hawks are characterized by their strong, hooked bills and strong talons. There is a widespread belief that all hawks are natural enemies of man because they sometimes kill songbirds, game birds, and mammals. Instead, hawks are almost entirely beneficial, and are of considerable value in the control of rodent populations and in killing sick and unfit animals. The predation of hawks on desirable species is generally limited to surplus animals that are too numerous for the food supply. Most species of hawks are now given legal protection, and several states protect all hawks.



(a) Red-shouldered hawk, *Buteo lineatus*; length 23 in.; (b) Sharp-shinned hawk, *Accipiter striatus*; length to 12 in (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

The voice of hawks is loud and harsh. They build a rough, bulky nest of sticks, usually high in a tree; some species nest on the ground, however. The young are covered with a sparse coat of down which grows to a thick coat in a few days. Hawks strike their prey with their feet, and tear it into chunks with their bills.

There are three families of hawks. The family Pandionidae has only one species, the osprey, *Pandion haliaetus*. This bird is often called the fish hawk, because fishes are its only food. The remainder of the hawks may be roughly divided into the pointed-winged hawks, or falcons, of the family Falconidae, and the broad-winged hawks of the family Accipitridae. The latter may be further distinguished as the buteos, with broad wings and short, broad tails, and the accipiters with short, rounded wings, and long tails. Representatives of these are shown in the illustration. *See* FALCONIFORMES.

[J.D.B.]

Hay fever

An allergic disorder of the nose and related structures due to sensitization by certain plant pollens. The term hay fever, although popularly used, is incorrect since hay is rarely implicated and fever is seldom present. There is continuous or intermittant nasal obstruction accompanied by wheezing, sneezing, coughing, and the copious production of a watery discharge. The nasal mucosa becomes swollen and pale from vascular congestion and edema. Intense itching of the nose, mouth, and pharynx may occur, as well as excessive lacrimation, photophobia (light sensitivity), headache, irritability, insomnia, and gastric disturbances.

Hay fever, or allergic rhinitis, is generally precipitated by exposure to pollens. There are various peaks of incidence in different areas, depending on the existing vegetation. In the United States, spring, summer, and fall types of hay fever are related to the pollens of certain trees, grasses, and ragweed, respectively.

Skin testing and desensitization procedures are often helpful in identification of the inciting agent and in decreasing the severity of the reaction. Related forms of nasal allergy may be caused by exposure to other allergens, such as foods, drugs, contactants, physical agents, and emotional crises in susceptible individuals. These may occur only in the nasal region but often are seen as part of a more generalized body reaction. *See* HYPERSENSITIVITY.

[E.C.ST.]

Head

The region of the body consisting of the skull, its contents, and related structures. Its two principal parts are the cranium or brain case and the face.

The skin, hair, and subcutaneous tissues over the top of the skull are collectively known as the scalp. The regions of the skull take their names from the underlying bones, for example, the temporal, parietal, frontal, and occipital regions.

The intracranial contents include the brain and uppermost portion of the spinal cord with their coverings, blood vessels, and the important cranial nerves, as well as the cerebrospinal fluid system. Many openings, or foramina, afford means of passage from within the skull for nerves and blood vessels.

The heads of all vertebrates are grossly similar; they vary in size, proportions, developmental state, and details of structure, but not in basic plan. *See* CRANIUM; NOSE.

[E.C.ST.]

Headache

A deep, as opposed to superficial or cutaneous, form of pain partaking of the aching quality of other deep pains and distinguished from them chiefly on the basis of locus of origin (*see* PAIN, DEEP). Most headaches may be shown to derive from distending forces applied to intracranial blood vessels. The presumption is, therefore, that the nervous structures responsible for initiating the massive

patterns of sensation involved in headache are the free nerve endings embedded in the walls of vascular structures inside the head. Stretching of these fibers by agents producing traction, displacement, distention, or inflammation constitutes the stimulus, even as such forces, when exerted on visceral organs and muscle tendons, evoke deep aching pain and, when acting on superficial cutaneous tissues, arouse prickling or burning pain.

In the search for the underlying causes of headache, some possibilities may be quickly eliminated. The brain itself, namely, its parenchymatous tissue, does not give pain when cut, pinched, or subjected to electrical stimulation. Moreover, most of the brain covering is similarly insensitive. However, the structures that anchor the brain mass to the cranium, intracranial veins chiefly, and the important blood vessels that provide nourishment to the brain, the arteries feeding all parts of the brain mass, are freely supplied with pain endings.

[illegible]

the total cerebrospinal fluid (about 20 ml) is withdrawn from the spinal canal by lumbar puncture with the subject in an erect position, a violent headache may be created. If the pressure is restored with injection of an equal volume of physiological saline solution, the pain is allayed. If the bodily position is changed to horizontal, or the head is moved, the aching pain may be reduced. Conversely, greatly increasing the intracranial pressure does *not* initiate headache. Indeed, this tends to stabilize the cranial structures and reduce the pounding or pulsating of the vascular tissues.

Among the most intense headaches are those associated with migraine, severe throbbing pains localized chiefly in the frontal [illegible]

[illegible] of central origin. Migraine headaches may be induced by any of a number of toxic agents, by body-temperature changes (artificial fever), and by injection of histamine. All of these have been used experimentally. All distend cerebral blood vessels; all may be counteracted by increased intracranial pressure.

Headaches may be induced by any set of conditions that will create and sustain strong contractile states of muscles in the head or neck region. This appears to be the basis of chronic headache from eyestrain. Any procedure eliminating the cause of visual discomfort, such as adequate ocular correction, can be counted on to diminish or obliterate the muscle tension and thus the primary cause of pain. Similarly, headache from hypertension is likely to be an indirect effect, one owing its immediate causation to prolonged tension of head muscles.

The persistent and severe headaches associated with many brain tumors are traceable to the steady traction imposed on vascular structures by mass displacement of tissue. Unless there has been a long history of development of the tumor, in which case the headache may become generalized, the localization of the head pain is likely to be of diagnostic importance, revealing with some fidelity the locus of the tumor. [F.A.G.]

Bibliography: H. G. Wolff and S. Wolf, *Pain*, 1948.

Health physics

The science that deals with problems of protection from the hazards of ionizing radiation or prevention of damage from exposure to this radiation. It is a border field of physics, biology, chemistry, medicine, engineering, and industrial hygiene. Health physics is concerned with radiation protection problems involving research, engineering, education, and applied activities. It deals with mechanisms of radiation damage, methods of measuring and assessing the radiation dose, devices for reducing or preventing radiation exposure, and the establishment of maximum permissible exposure levels. Health physics had its beginning in 1942 along with the nuclear energy and reactor programs at the University of Chicago. It is estimated there are over 2000 practicing health physicists throughout the world. An international organization, the Health Physics Society, was formed in 1956, and two years later it had a membership of about 1000. *See* DECONTAMINATION (RADIOACTIVE CONTAMINANTS); MONITORING (IONIZING RADIATION); RADIATION DAMAGE (INANIMATE MATERIALS); RADIATION INJURY (BIOLOGY); RADIOACTIVE WASTE DISPOSAL; RADIOISOTOPE PRODUCTION. [K.Z.M.]

Bibliography: H. Etherington (ed.), *Nuclear Engineering Handbook*, 1958.

Hearing

The general perceptual behavior and the specific responses made in relation to sound stimuli. The perceptual behavior of hearing has many dimensions, including sound arousal, perceptual orientation, sound detection, sound discrimination, and perceptually motivated listening, such as music appreciation. Hearing behavior also involves the internal biochemical and physiological states associated with emotion, which add esthetic value and emotional tone to perception of speech, music, and noise. The science of audition has been concerned principally with analysis of auditory detection and discrimination as related both to the properties of the sound stimulus and to physiological events in the ear and in parts of the nervous system. Knowledge of other aspects and behaviors of hearing is quite limited. For responses of infrahuman forms see PHONORECEPTION. *See also* PSYCHOACOUSTICS.

Perceptual activities. The perceptual activities in hearing are diagrammed in Fig. 1 as a series of vital interactions between the physical energy of sound, the processes of the ear, the neural systems of the brain, and related neurochemical and neuromotor activities. Five forms of response which alter the metabolic or behavioral state in relation to sound stimuli are skeletal muscle reaction, visceral muscular response, peripheral glandular response,

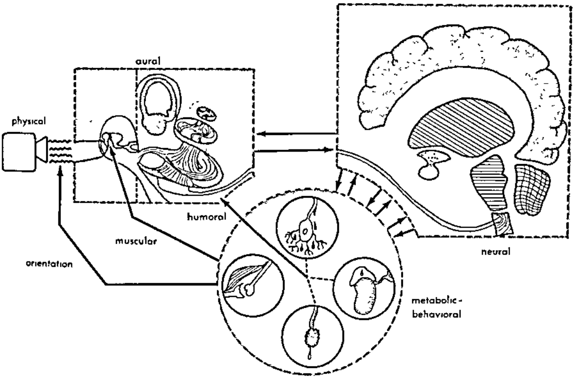


Fig. 1. Physical, aural, neural, and metabolic-behavioral levels of interaction involved in hearing.

endocrine activity, and neurohormonal action involving release of such compounds as adrenalin, acetylcholine, and serotonin in synaptic centers. Figure 1 shows the feedback relationships involved in integrated hearing behavior. Muscular activity, as in arousal behavior, changes the posture and state of the body in relation to the sound source. Contractions of small muscles in the middle ear alter sound transmission by the eardrum. Visceral activity may supply more blood to the ear and brain, thus changing the oxygen tension of the system and the sensitivity of the organism to sound stimuli. Hormones present in the nervous system both facilitate and inhibit synaptic activity, thus changing the state of the organism, possibly of the ear itself. Release of chemical substances by secretory glands, such as tears and perspiration, may alter the efficiency of auditory discrimination as well as perceptually motivated activities such as sustained listening.

Hearing is thus not a single process, but a generalized activity involving several dimensions of response, each of which serves certain integrative functions in maintaining the efficiency of the individual in responding to sound stimuli. These phases of hearing behavior differ especially in their timing relative to the onset of the sound stimulus.

Sound arousal. Sound arousal is attracting attention by sound stimulation. Hearing, along with cutaneous perception and visual perception, is a significant means of arousal to action. Sounds can arouse people from deep sleep and activate the individual under many other circumstances.

The arousal reaction by sound or other stimuli is a release or triggered type of response involving reflex changes in posture, as well as selective inhibition and activation of smooth muscles and glands. The pupils of the eye dilate and there are reactions of the small muscles of the middle ear. The reaction time of this startle pattern to sound stimuli may be as low as 0.01–0.05 sec.

Orienting reactions. Orienting reactions are next in time. In man these consist of turning the head and assuming postures which facilitate the reception of the sound stimuli in question and reduce the effectiveness of interfering sounds. Orientation to sound is much more obvious in many animals, involving specific movements of the ears as well as more general postural responses.

Sound detection This consists in finding or noting a sound source within a background environment of other sounds or other kinds of stimuli.

Sound discrimination. This involves comparing, ordering, or arranging two or more sounds in terms of the discriminable properties of sound: pitch, loudness, timbre, volume, timing, duration, and location in space.

Motivated listening. Motivated listening involves all of these primary phases of hearing integrated into sequential motivated behavior in response to sustained speech, music, or noise. It is sound perception organized in relation to adaptive behavior.

Listening can be measured in terms of the stimulus conditions of its occurrence and by means of psychological scales. Administration of attitude scales and preference scales, measuring the time of listening to sustained speech or music, and recording the level of emotion and motivation are all means of assessing the perceptual value of music or speech sounds. A critical dimension of listeni

is the semantic value of symbols, words, code signals, or other sounds. Standardized articulation tests made up of syllables, words, or short sentences give a measure of the effectiveness of speakers and of sound generating and transmitting devices in providing meaningful information to a listener.

Listening may be directed toward specific sounds or it may be supportive. The presence of some effective sound in the environment appears to be essential for the psychological well being of most individuals. The disorganizing effects of isolation on man and animals appear to be related in part to the supportive value of sound and of listening, for complete and prolonged absence of sound has been found to cause marked emotional depression. Music can serve a positive supportive function in work situations and can facilitate the rhythm of many work and play activities. Even moderate levels of noise may act as an alerting stimulus condition and facilitate work. Excessive sound, however, either music or noise, acts as a stress source, and can disturb efficiency of certain forms of work behavior. In addition to direct injury to the ear, excessive noise may also produce long-term stress effects on the auditory mechanism as well as the whole neurophysiological system.

Although rudimentary responses to sound can be observed in some fishes, amphibians, and reptiles, only birds and mammals, with their highly developed organs for the reception of sound stimuli, can be said to possess refined hearing. To understand the nature and significance of sound perception in these animal forms, one must understand both the integrated behavioral activities in such perception and the functions which this behavior serves in evolutionary development. In some mammals, such as the bat and the porpoise, hearing is especially refined to serve as a kind of radar in echo location and ranging. Man's sensitivity to sound is closely correlated with the physical characteristics of the sounds he produces in speech. Comparable examples could be multiplied to show that the nature and efficiency of hearing in animal species reflect the development of sound-controlled behavior by processes of natural selection.

Sound properties. The properties of sound can be described in terms of the physical correlates of the sound stimulus, and in terms of the psychological dimensions of auditory response. The three main physical properties of sound are frequency, intensity, and complexity of sound waves. The principal psychological properties are perceptions of pitch, loudness, and timbre. Although pitch is closely related to frequency, loudness to intensity, and timbre to complexity, there is no one-to-one relation between the physical and psychological properties of sound. For example, the loudness of a perceived sound is affected by the frequency of a sound stimulus as well as its intensity.

Sounds originate from a mechanical displacement of an elastic body, such as a tuning fork or taut string. When such a body is made to vibrate,

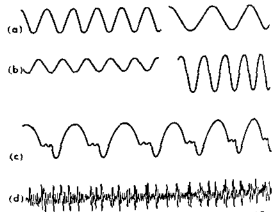


Fig. 2. Stimulus correlates of perceived sound. (a) Two tones differing in frequency. (b) Two tones differing in intensity. (c) Complex musical tone. (d) Noise

it imparts some of its energy to the surrounding air (or other medium), thus causing air particles to vibrate in a pattern like that of the source. A series of alternate regions of condensation and rarefaction of the air particles is established, which proceeds outward from the source of the disturbance. It is the progressive movement through the air of the pattern of changing pressures from the vibrating source that is called the sound wave.

The waveforms of sound may be either pure or complex (Fig. 2). Pure sounds occur rarely in nature, but can be produced under controlled laboratory conditions with special equipment. The sounds of everyday life are complex.

In its simplest form, a sound wave is described as simple harmonic motion, portrayed visually as a sine curve (Fig. 2*a*, *b*). One upward swing of the curve and one downward swing represent one rarefaction and one condensation, or a single cycle of vibration. The number of complete cycles occurring each second determines the frequency of the sound wave, while the square of the amplitude of displacement or height of the curve represents the intensity of the sound. The two waves in Fig. 2*a* represent pure tones of the same intensity but with different frequencies, while the two waves in Fig. 2*b* represent tones of the same frequency and different intensities.

Sound waves can also vary in complexity. Figure 2*c* represents the waveform of a musical tone. It can be seen that the pattern repeats itself, or is periodic. Figure 2*d* represents noise, a complex sound stimulus with no apparent periodicity in its wave form.

A complex wave can be described as being made up of a number of different sine waves, each of which has a definite frequency and intensity. In musical tones, the frequencies of these components are multiples of the lowest frequency represented. Thus a tone might include component frequencies of 50, 100, 150, and 200 cycles per second (cps), each of a certain intensity. The lowest frequency is called the fundamental, and the others overtones, or harmonics. Different musical instru-

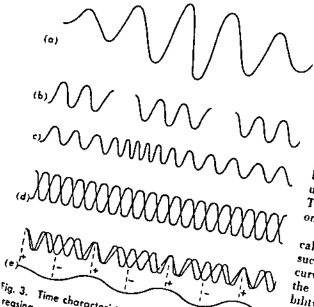


Fig. 3. Time characteristics of perceived sound (a) Increasing and decreasing loudness. (b) Interrupted sound. (c) Warble. (d) Phase cancellation. (e) Beats.

ments produce distinctive harmonic patterns. It is these differences in timbre which enable man, for example, to differentiate between middle C as played on the piano and as played on the oboe. Noise is a more or less random assortment of frequencies with no recognizable pitch. A complex sound which includes all audible frequencies in equal energy ratios is called white noise.

Some auditory effects are due to the critical time characteristics of the sound wave patterns. A wave which gradually increases and decreases in amplitude is heard as a tone which waxes and wanes in loudness (Fig. 3*a*). An interrupted pure wave (Fig. 3*b*) may be heard as an interrupted sound with tonal value or, if the sound intervals are very brief, as a series of clicks. The frequency of a wave may be modulated to give a warble (Fig. 3*c*). Tones of the same frequency and amplitude may vary in phase from 0 to 360°; two such tones which are 180° out of phase (Fig. 3*d*) cancel each other out so that no sound is heard. When two tones nearly equal in frequency are sounded together (Fig. 3*e*), the perceived effect is of one tone which beats, or waxes and wanes in loudness. The number of beats per second is equal to the frequency difference between the two tones; this difference determines the number of times per second that the tones coincide in phase and reinforce each other.

When two tones differing in frequency around 30 cps or more are sounded together, one may hear not only beats, but also tones not actually emitted by the sound sources. For example, if two tones with frequencies of 2000 cps and 2100 cps are sounded together, one may hear a difference tone corresponding to the difference of 100 cps, and, under some conditions, a summation tone corresponding to the sum of the two frequencies, 4100 cps. These combination tones are produced by processes within the auditory mechanism. *See* SOUND.

Sound detection. There are two aspects of sound detection which should be considered: detection of a sound stimulus in an otherwise sound-free field, and detection of a sound stimulus in the presence of a second masking sound.

Limits of hearing. The frequency and intensity patterns which are perceived as sound have definite limits, which vary to some extent from one individual to another and under different conditions. These limits of hearing also vary markedly from one animal species to another.

The limits of hearing are represented graphically as the area of hearing (Fig. 4). To obtain such a chart, two curves are plotted, the audibility curve, which is the lower curve in the figure, and the upper curve of "pain" or "feeling." The audibility curve is determined by measuring the absolute threshold for a series of frequencies within the hearing range. The absolute threshold for a given frequency is the lowest intensity at which that frequency can be heard. The upper limit of the area of hearing is determined by increasing the intensities of tones until they produce a perception of pain, tickle, or feeling.

Figure 4 shows that human individuals normally perceive as sound those frequencies between about 16 cps, or lower, and 20,000 cps, with the greatest sensitivity around 1000–4000 cps. That is, the audibility curve, showing absolute thresholds for dif-

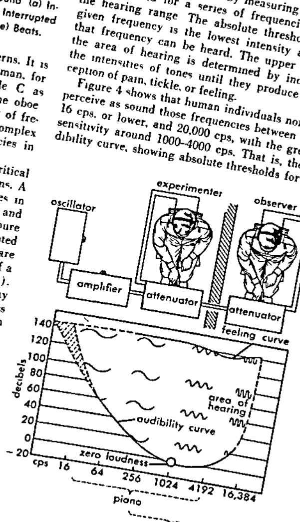


Fig 4. Measuring the limits of hearing. Absolute thresholds are determined by the method of adjustment.

ferent frequencies, dips lowest for frequencies of around 1000–4000 cps.

The range of sound intensities is expressed in terms of units called decibels, which are not absolute units of measurement, but express a ratio between a given sound intensity and a reference sound. A bel is the logarithm to the base 10 of the ratio of a sound of given intensity to the reference sound; a decibel, the more commonly used unit, is one-tenth of a bel. A convenient reference is an intensity level which approximates that of a tone of 1000 cps which can barely be heard by a normal individual. This standard is used as the zero point of the vertical intensity axis in Fig. 4. A sound 10 times as intense has a value of 1 bel or 10 decibels (db), and a sound 100 times as intense, 2 bels or 20 db Thus the upper limits of hearing in Fig. 4 indicate that, at some frequencies, man may respond to sounds of 140 db, or 10^{14} times as intense as man's absolute threshold at 1000 cps.

The lower curve in Fig. 4 represents the threshold at which observers hear something as distinct from absolute silence. When observers are told to report when they hear tones characterized by pitch, this tonality threshold is about 3–7 db higher than the audibility threshold.

The area of hearing varies with age as well as with injuries and other defects Sensitivity to higher frequencies typically decreases with age. Some animals, such as the bat, mouse, rat, cat, and dog, have a frequency range beyond that of man. The bat may hear tones as high as 75,000 cps. "Silent" dog whistles produce tones above the range of human hearing, but still within the dog's upper limit.

Auditory masking. It is sometimes impossible to detect one tone in the presence of another tone or noise The first tone is said to be masked by the second tone or noise. The masking effect is defined as the difference between the absolute threshold of the masked tone and the threshold of detection of this tone in the presence of the masking sound. In general, the masking effect is greater for tones near the masking sound in frequency, although tones higher in frequency than the masking sound are affected more than lower tones. The range of frequencies affected is greater, the greater the intensity of the masking sound, and the lower the frequency of the masking sound.

If speech sounds occur in the presence of noise, they may be completely masked. As the intensity level of the speech is raised, it may be detected as sound while it is still unintelligible. If the intensity is raised still further, the speech becomes intelligible. Thus there is a threshold of detectability and a threshold of intelligibility for masked speech. Either of these thresholds increases in a linear way as the intensity of the masking noise is raised above about 30–40 db. *See* MASKING (SOUND).

Sound discrimination. The differential threshold is a measure of ability to detect a difference between stimuli. The usual method for determining such thresholds for intensity and frequency of sounds is to introduce changes, either of intensity or frequency, into continuous sounds, and measure change the smallest that can be detected.

When white noise is used as a stimulus sound, the smallest change in intensity which can be detected remains fairly constant at slightly under 0.5 db, above an intensity level of about 20 db. Below 20 db, the just detectable change in intensity increases sharply. When pure tones are used as stimuli, a similar threshold of 0.5 db or less is found above levels of about 40 db for a 4000 cps tone. The threshold is greater at lower intensities, and for tones of lower and higher frequencies. Since the decibel is a relative measure of intensity, one can say that the relative change in intensity which can be detected remains fairly constant above 20 db for noise, and at somewhat higher intensity levels for pure tones.

The smallest change in frequency of pure tones which can be detected is fairly constant at about 2–3 cps for frequencies up to 1000 cps, and at intensity levels over about 20 db. Above 1000 cps and 40 db, the relative change in frequency which can be detected remains fairly constant, at about 0.2–0.4%.

In general, loudness increases with increased intensity, but is also a function of frequency. The audibility threshold in Fig. 4 is an equal loudness contour of tones that can just barely be heard At different frequencies such equal loudness tones require different intensities. Similar contours can be plotted at different intensity levels by asking observers to adjust tones of various frequencies to be equal in loudness to a standard. Loudness depends to some extent on duration of the sound stimulus. Up to about 2000 milliseconds (msec), the longer the duration of a tone, the lower its threshold.

Pitch varies mainly with the frequency of sounds, but also with intensity. As the intensity of high-pure tones is increased, the apparent pitch of high-frequency tones gets higher, and that of low-frequency tones gets lower. Perception of pitch also depends on a certain minimum duration of the tonal stimulus. Tones shorter than about 5 msec are heard as clicks.

Other properties of tones which have been suggested as discriminable qualities are volume, or spatial voluminousness; density, or compactness; and brightness. Lower tones appear to have more volume than higher tones of the same intensity, while higher tones appear to have greater density.

Binaural hearing. Special characteristics of sound discrimination arise from binaural stimulation, or stimulation of the two ears.

Binaural localization of sound. The efficiency with which sound sources are localized in the environment is dependent on both learned cues and the ability of the auditory system to respond to slight differences in stimulation of the two ears. Success in judging distance depends on recognition of the sound, knowledge of how such a sound

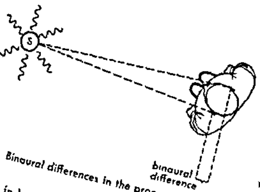


Fig. 5. Binaural differences in the properties of sound.

changes in loudness and timbre with distance, and corroboration from other sensory cues, especially visual. Even when other cues are eliminated, the direction of a sound which stimulates both ears can be judged fairly accurately except when it is exactly behind, above, or in front of the observer. Sounds to the right or left are localized in terms of differences at the two ears of intensity, time of arrival, or phase (Fig. 5). Kinesthetic cues from head movements combine with binaural cues to enable observers to judge the elevation of sounds. When pure tones are used as stimuli, it is found differences in phase are effective cues for localization for low tones, up to about 1500 or 2000 cps. High tones can be localized in terms of differences in intensity at the two ears, but lower frequency tones reach the ears with very little intensity difference. At frequencies near 3000 cps neither type of cue is effective, and localization is very poor. Complex sounds are ordinarily localized more efficiently than tones because both temporal and intensitive cues are operative. Sounds of short duration are judged in terms of time of arrival. The auditory system contains neural mechanisms which are especially sensitive to differences in time of arrival or phase of two sounds.

Binaural beats are heard when tones differing slightly in frequency are presented separately to the two ears.

Binaural summation. The ability to detect and discriminate sounds is increased with binaural stimulation. Tones heard with both ears sound louder than when heard with one. This summation effect accounts for the somewhat lower absolute thresholds found for binaural stimulation as opposed to monaural, and also for the lower intensity discrimination thresholds which are found for the two ears. Frequency discrimination is also better with binaural stimulation, especially for tones of very low and very high frequencies.

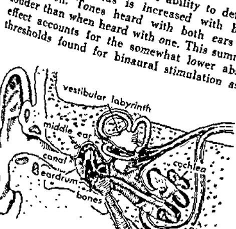


Fig 6. Diagram of the structure of the ear.

Response of auditory system in hearing. The ear (Fig. 6) is a system for transducing the mechanical energy of sound into electrical energy which triggers nerve impulses in the auditory nerve. As sound strikes the eardrum it causes the three small bones of the middle ear to vibrate. These movements set up corresponding vibrations in the oval window, and thence in the fluids of the cochlea.

The auditory organ of the inner ear, the cochlea (Fig. 7*a*), is an enclosed structure which is coiled upon itself for 2½ turns. It contains three fluid-filled canals: the ascending (vestibular) canal, which leads from the oval window to the apex (helicotrema); the descending (tympanic) canal, from the apex down to the round window; and between

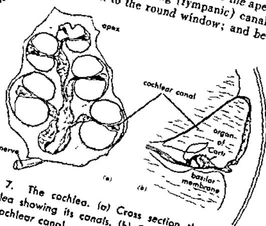


Fig. 7. The cochlea. (a) Cross section through the cochlea showing its canals. (b) Cross section through the cochlear canal.

them the smaller cochlear canal (Fig. 7*b*). The gelatinous-fibrous membrane making up the base of the cochlear canal, the basilar membrane, is relatively narrow near the oval window and widens as it ascends toward the apex. Upon this membrane is the organ of Corti and its hair cells, which are thought to be the true receptors of hearing. Proliferations from the fibers of the auditory nerve extend up the center of the cochlea and connect with these hair cells.

Cochlear resolution of sound. The ability to hear many different frequencies as distinct pitches is related to the ability of the cochlea to resolve these frequencies. Between about 200 and 2000 cps, this resolution is accomplished by differential response of the basilar membrane. The cochlea has different resonance values at different points along its length, so that high tones cause the fluids and membrane to vibrate near the base, and low tones, near the apex (Fig. 8*a*). Vibrations of the oval window probably set up traveling waves in the fluids and membrane which cause maximal vibrations at the place with a natural resonance value corresponding to the sound frequency. Below 200

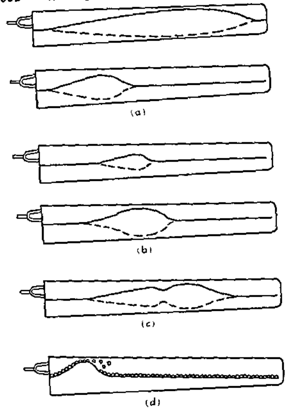


Fig 8 Basilar membrane effects. (a) Resolution of different frequencies by localized vibrations. (b) Representation of intensities by different amplitudes of vibration. (c) Masking caused by overlapping of vibrations (d) Destruction of hair cells by intense sound.

cps, the membrane vibrates as a whole, or nearly so. Different intensities cause vibrations of the membrane of different amplitude (Fig. 8*b*). If two tones are present, the vibration of the stronger will overlap the weaker, so that the latter is masked (Fig. 8*c*). A very intense sound may cause the hair cells to be thrown off the membrane at a place corresponding to the frequency level of the sound (Fig. 8*d*).

Cochlear potential and aural microphonic. The cochlea transduces sound energy of all audible frequencies into electrical energy. If electrodes are inserted into an animal's ear (Fig. 9), a potential of 40–100 microvolts is found to exist between the fluid of the cochlear canal and the organ of Corti, with the greatest potential near the base. This current, which exists when no sound stimulus is present, can be recorded by means of amplifiers and meters attached to the electrodes.

A sound stimulus acting on the cochlea modulates the sustained potential, setting up a cochlear or aural microphonic, which reproduces the sound frequency with high fidelity. This aural microphonic, which varies between 0.5 and 100 microvolts, can be amplified and recorded on an oscilloscope or reconverted to sound by means of a loudspeaker. The magnitude of these potentials varies in a linear way with the sound intensity. When different levels of the cochlea are tested, a maximal microphonic effect for low tones is picked up at the apex, and a maximal effect for high tones, at the base. The lower intensity thresholds for the microphonic response define a curve which is similar in form to the audibility curve for hearing. It is generally believed that the aural microphonic is a [illegible] either the hair [illegible] ers.

originating at the hair cells at all points along the organ of Corti, emerge from the cochlea to form the auditory nerve (VIIIth cranial nerve). This nerve enters the cochlear nucleus in the medulla oblongata, from which point connections are made with other centers in the midbrain, cerebellum, thalamus, and temporal lobe of the cortex (Fig 10). Discrimination of sound stimuli depends on differentiation of neural action at all levels Three principal kinds of differential neural response can be demonstrated.

In the first place, it can be shown that different regions of the basilar membrane have specific spatial representation in the auditory nerve and brain

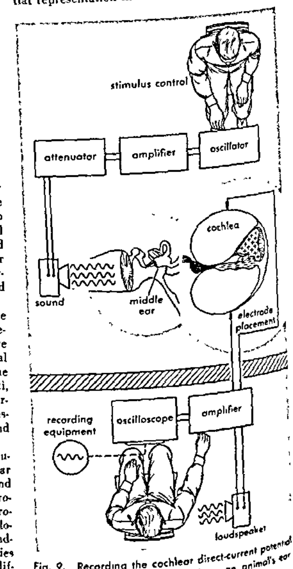


Fig. 9. Recording the cochlear direct-current potential and the cochlear microphonic from an animal's ear

centers. When the ear is stimulated by tones or clicks or when the cochlea is stimulated electrically (Fig. 10), the electrical response of the nerve, cochlear nucleus, midbrain, thalamus, cerebellum, or cortex can be amplified and recorded. By such means it has been found that the basilar membrane is projected spatially in all of these neural systems.

The way in which the response of the basilar membrane is represented in the auditory nerve is diagrammed in Fig. 11. Low-frequency sounds which cause the whole membrane to vibrate produce spatially scattered nerve impulses throughout the nerve (Fig. 11*a*). A sound between about 200 and 2000 cps which causes a localized vibration of the membrane sets up impulses in a specific group of nerve fibers located in a particular area

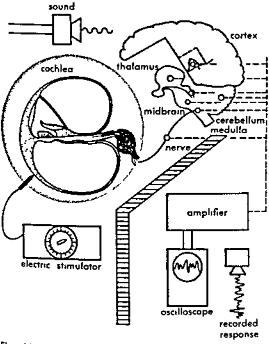


Fig 10. Recording the electrical responses of the nerve, cochlear nucleus, midbrain, thalamus, cerebellum, and cortex to sounds of different frequency and to electrical stimuli applied to different regions of the cochlea.

(Fig. 11*b*). This geographic representation of the different parts of the basilar membrane is also found in the auditory centers of the brain.

The second type of neural differentiation is a temporal one. Within limits, the frequency of a stimulus sound is reproduced by the synchronized electrical response of the auditory neural systems. This neural response does not have the fidelity of the cochlear microphonic. The electrical response of the total auditory nerve, for example, faithfully reproduces the sound stimulus to frequencies up to 4000–5000 cps, but not higher ones. At higher frequencies, the neural synchronization is very poor. Furthermore, the fidelity with which the electrical response of higher centers of the brain, such as

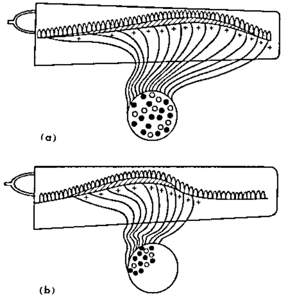


Fig. 11 Spatial representation of basilar membrane response in the auditory nerve (a) No spatial localization of low tone response. (b) Specific localization of response to frequencies from about 200 to 2000 cps.

the thalamus, cerebellum, and cortex, follows the sound stimulus in frequency is far less than that of the auditory nerve and lower brain centers. High-fidelity frequency differentiation is probably mediated by the lower auditory centers of the brain.

The ability of the auditory nerve and brain tracts to follow the frequency of sound cannot be based on the response of single neurons, for the maximal rate of firing of single auditory neurons is probably not over 700–800 impulses per second. The synchronization produced is explained by the volley principle (Fig. 12), which postulates that different single nerve fibers react alternately to successive waves in the sound stimulus. When the frequency of a sound is too high for any one fiber to fire at the same frequency, a number of fibers coordinate

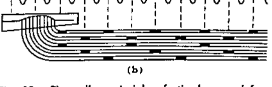


Fig. 12. The volley principle of stimulus-neural frequency synchronization. (a) Synchronization for a weak tone. (b) Synchronization for an intense tone.

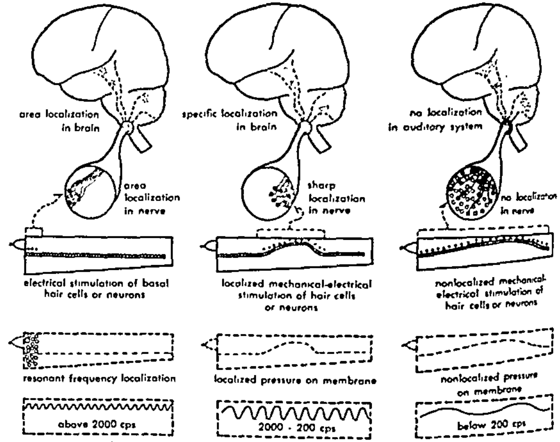


Fig. 13. Differential space-time patterns of cochlea, basilar membrane, auditory nerve, and central nervous system to sounds of different frequency.

their response to produce a synchronized response in the group of fibers which faithfully reproduces the sound frequency. If the intensity of the stimulus is increased, the total number of impulses in the nerve is increased but they still synchronize with the sound frequency.

Still another differential response has been found in second-order auditory neurons, one synapse removed from the neurons leading from the cochlea. By inserting electrodes of microscopic size into the cochlear nucleus, it is possible to record the electrical response of single second-order neurons. Neurons can thus be found which are selectively tuned to certain frequencies, at which frequency less energy is required to excite them. Although this neuron tuning is very sharp near threshold, as the sound intensity is increased, each neuron responds to a broader band of frequencies, typically spreading more toward lower frequencies than toward higher. Neurons responding to very high frequencies apparently are more sharply tuned than those responding to lower frequencies.

Space-time pattern theory. The various facts about cochlear and neural response make possible a theoretical interpretation of how sounds of various frequencies and intensities are discriminated (Fig. 13). The evidence indicates that pitch discrimination at different frequency levels depends either mechanical or neural analysis or a combination of both. Sounds below about 200 cps cause the basilar membrane to vibrate almost as a whole; the activity set up in the hair cells and auditory nerve fibers has no specific (or but poorly defined) localization. Perceived pitches of such low frequencies depend on the over-all synchronization of nerve impulses with the frequency of vibration of the whole membrane.

Above about 200–300 cps, localization of membrane vibrations becomes much sharper, as does the spatial representation in the nerve and brain centers. The pitch of these middle frequencies up to about 2000 cps may be defined by the particular place of stimulation, but may also depend on synchronized neural response frequencies. Pitches corresponding to this middle range of frequencies can be produced by introducing white noise to the two ears with a constant phase difference at a certain frequency level within the noise. Such a binaural tonal effect could only be based on the temporal relationship of neural impulses in the brain.

The critical basis for discriminating frequencies above about 2000 cps is unclear. The best interpretation is that they activate the fluids of the cochlea but cause no vibration of the membrane except at their onset and termination. It is assumed that these higher frequencies are localized in a general way at the basal end of the cochlea because of the resonant characteristics of the cochlear chambers.

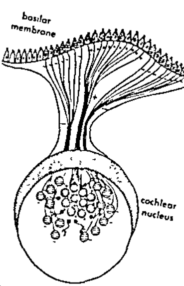


Fig. 14. The principle of neural funneling, that is, tonal sharpening and isolation, by neural summation and inhibition in the cochlear nucleus.

and because of the relatively greater strength of the cochlear action potential at the basal end. The observations that specific neurons may be tuned to these high frequencies were made on second-order neurons in the cochlear nucleus. Similarly tuned first-order neurons are not known.

In general, the most widely accepted theory of hearing is multidimensional. Pitch perception is believed to depend upon a combination of place localization of mid-frequency pitches in the cochlea and nervous system, and of frequency representation, especially of low and high pitches, by means of synchronized neural response.

Auditory funneling. The neural response in hearing involves both excitation and inhibition. That is, the neural effects of each stimulus include both activation of critical neurons in the auditory pathway related to certain tones and the inhibition of other pathways associated with other tones. Thus the system can isolate tones of specific pitch and loudness, and attend to and localize these tones.

One stage of neural funneling occurs in the cochlear nucleus by means of the activity of a band of efferent nerve fibers which extend from the nucleus back into the organ of Corti. When the vibrations of the basilar membrane are localized (Fig. 14), hair cells at the maximum point of vibration (black) are more intensely stimulated than adjacent cells (striped). In the cochlear nucleus the more intensely excited neurons activate surrounding efferent neurons, which send impulses back to the organ of Corti to inhibit the action of the adjacent fibers.

Integrative funneling through summation and inhibition of specific neuron pathways takes place at all levels. It is the process involved in suppression of the electrical activity of the cochlear nucleus when the individual is aroused by visual or tactual stimuli. Ability to perceive separate tones in a complex of tones independently of their phase relations (Ohm's acoustic law) is a function of this process of neural sharpening of the effects of sound on the cochlea.

Integrative control. Integrative control involves both neural and biochemical pathways in determining the muscular, neurohumoral, humoral, organic, and metabolic activities of hearing, and processing the feedback information from these activities. The levels of interaction involved in various aspects of hearing behavior are detailed in Fig. 15. The properties of sound perception, such as pitch and loudness, which correlate closely with variation in the sound stimulus are determined very largely by the functions of the cochlea itself. Other effects, such as combination tones, are defined by processes of distortion in both the cochlea and in nerve transmission. Still other properties are dependent on the integrative processes of funneling. Finally, more complex levels of integration determine emotional effects, consonance, disso-

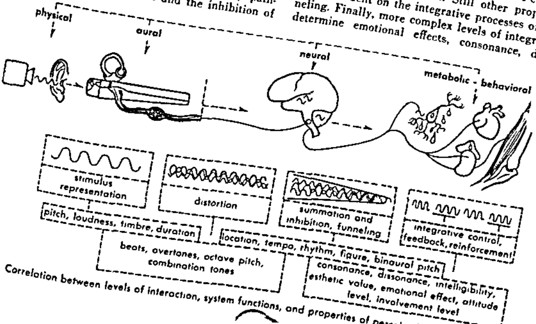


Fig. 15. Correlation between levels of interaction, system functions, and properties of perceived sound

nance, intelligibility, semantic value, and specific response control and reinforcement related to learning to perceive sound and to regulate motivated listening.

The different aspects of hearing behavior are affected by all types of integrative control. Sound discrimination can be made more precise by permitting the observer to regulate the source of sound. Such emotional effects of sound as consonance and dissonance can be understood only by taking into account the nature of humoral and neurohumoral responses of the listening individual. Learning to use sounds in a meaningful way for communication involves the neurohumoral and humoral systems as well as specific patterned psychomotor activities. The auditory mechanism is a unified system in which high fidelity of reception of sound, precise articulated control of sound sources, and generalized emotional and psychomotor adaptation to sound are not only possible but characteristic. See AUDIOMETRY; DEAFNESS; LOUDNESS; PITCH; PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL. [K.U.S.]

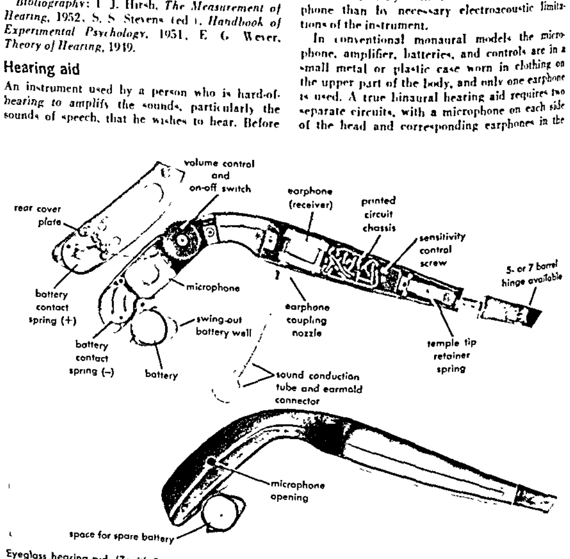
Bibliography: I. J. Hirsh, *The Measurement of Hearing*, 1952; S. S. Stevens (ed.), *Handbook of Experimental Psychology*, 1951; E. G. Wever, *Theory of Hearing*, 1949.

Hearing aid

An instrument used by a person who is hard-of-hearing to amplify the sounds, particularly the sounds of speech, that he wishes to hear. Before about 1920 the best hearing aids were ear trumpets and speaking tubes. The speaking tube, into which the talker speaks directly, is still a very simple and effective device with a single talker, but it is not very convenient. Modern electrical hearing aids are small enough to be worn comfortably and inconspicuously in the clothing, in the hair, behind the ear, or in the frames of spectacles. The essential components are a microphone, an amplifier (now almost always of the transistor type in wearable aids), a small earphone, and a plastic insert earmold which supports the earphone. Sometimes a vibrator is held by a spring headband behind the ear and delivers sound to the ear by bone conduction.

Many hearing aids have tone controls to give greater emphasis to high or to low tones, and practically all have adjustable gain controls. The frequency range of effective amplification usually extends from about 400 or 600 to about 3000 cycles per second. The useful output of a hearing aid is more often limited by the tolerance of the user for loud sound or by acoustic feedback from the earphone than by necessary electroacoustic limitations of the instrument.

In conventional monaural models the microphone, amplifier, batteries, and controls are in a small metal or plastic case worn in clothing on the upper part of the body, and only one earphone is used. A true binaural hearing aid requires two separate circuits, with a microphone on each side of the head and corresponding earphones in the



Eyeglass hearing aid (Zenith Radio Corporation)

two ears. There are many intermediate types and arrangements. The lower temple bar shown in the illustration is a dummy bar in the case of monaural hearing, but contains equipment identical to that in the upper temple bar for binaural hearing.

The group hearing aid is widely used in schools for the deaf as a permanent classroom installation. The teacher wears the microphone, or it stands on her desk. A single amplifier is operated from the power lines. Each student wears a pair of earphones and plugs in at convenient locations at desks or blackboards. Individual gain controls are located at the outlets. A desk type of hearing aid, suitable for office use, also operates from power lines instead of batteries. See HEARING, see also DEAFNESS. [H.DA.]

Bibliography: H. Davis and R. Silverman (eds.), *Hearing and Deafness*, 2d ed., 1960.

Heart

The muscular pumping organ of the cardiovascular system. In man, the heart weighs about 300 grams and is located behind the breastbone and ribs between the third to fifth costal cartilages. Its upper portion, or base, is directed to the right and backward and is the area where the great vessels enter and leave the heart. The lower muscular portion ends in a blunt apex which lies behind the fifth costal cartilage on the left.

The fibromuscular septum divides the heart into two parts, each consisting of a thin-walled receiving chamber or auricle, and a thicker, muscular pumping chamber, or ventricle. Blood enters the right auricle from the superior and inferior vena cavae which drain most of the body. It passes through the tricuspid valve to the right ventricle and is pumped to the lungs during systole, or contraction of the heart. Blood returns from the lungs by way of the pulmonary veins to the left auricle, passes into the left ventricle through the mitral valve, and during contraction is pumped out into the aorta.

The muscular wall of the heart, the *myocardium*, is lined with an inner endocardium and is covered externally by membranous pericardium. Additional structures include the coronary arteries and veins,

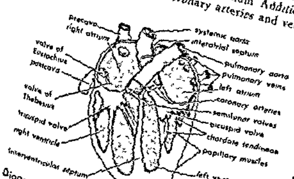


Diagram showing internal structure of four-chambered mammalian heart, ventral view. (From C. K. Weichert, *Anatomy of the Chordates*, 2d ed., McGraw-Hill, 1958)

the neuromuscular conducting system, autonomic nerve supply, and the remnants of the prenatal circulation, such as the fossa ovalis. See CARDIOVASCULAR SYSTEM. [E.C.ST.]

Heart (ganglionic pacemaker)

The only hearts definitely known to be paced by nervous discharge instead of rhythmic muscle fibers are those of the xiphosuran *Limulus* and the higher crustaceans. A. J. Carlson's classical work on *Limulus* established the neurogenic origin of the heart beat in the distinct threadlike ganglion in the dorsal midline of the heart. Considerable work on this preparation and on the hearts of crayfish, lobsters, and crabs has led to the following picture.

The striated muscle fibers of the myocardium are tetanically driven by a burst of nerve impulses for each beat. Unlike the all-or-none contraction of vertebrate hearts, they respond with graded and prolonged rather than twitchlike contractions. Individual muscle fibers probably often give a bumpy, unfused tetanic electrogram and mechanogram and perhaps receive several nerve fibers.

The form of the electroneurogram, electromyogram, and mechanogram is reproduced again and again for hundreds of beats as a result of the consistency of the pattern of nerve impulses in the bursts. This pattern is made up of several to many impulses from each of several to many nerve cells, each firing time after time with the same sequence, frequency, and duration.

Nerve supply. The anatomy of the *Limulus* ganglion is not well known; it is large and many-celled. The largest cells are believed to be the normal pacemakers. Higher crustaceans have commonly from 9 to 15 nerve cells which include large and small cells, and monopolar but more generally bipolar and multipolar cells; details are mainly the result of work by J. S. Alexandrowicz. In the most-studied hearts, those of lobsters of several kinds, there are 9 cells of which 5 are large and anterior, 4 small and posterior; at least the former are in fairly constant positions. Complex connections among them form a neuropile with four or more kinds of synapses. The small cells project to the large cells and the latter to the muscle; other connections cannot be summarized here and are little known. A single extrinsic inhibitor nerve fiber and two accelerators come from the central nervous system, on each side, traveling in separate nerves part of the way; they distribute endings widely in the ganglion.

Physiological analysis. Physiological analysis, inaugurated by Maynard in lobsters, has indicated the large cells to be normally followers and motoneurons but with some interaction and different degrees of spontaneity, expressed if the higher frequency driving is removed. In the normal ganglion the pace of the heart is set by one of the small, posterior cells, this role possibly rotating among them. They do more than start the burst because they typically fire more impulses over a longer period than the large cells. Intracellular recording with ultramicroelectrodes confirms this distinct

and shows that followers have several types of synaptic potentials, pacemakers probably none. Synaptic potentials in response to impulses arriving from the pacemaker cell are large after the inter-burst interval and small after shorter intervals as during the burst; this may be called defacilitation. Synaptic potentials from impulses arriving in the accelerator nerve fibers are of two kinds: in some cells they resemble the foregoing but are initially small, increasing with repetition, and with higher frequency. This is facilitation. In other cells there is evoked only a maintained shift of the membrane potential in the depolarizing direction, without discrete humps for each arriving presynaptic impulse. The inhibitor nerve fiber, when stimulated outside the heart at adequate frequencies, causes synaptic potentials which facilitate and which are in some cells hyperpolarizing, in others depolarizing, decreasing and increasing the probability of an impulse arising, respectively.

The large follower cells never depolarize completely or explosively, that is, impulses arise in their processes and do not invade the cell body. Spikes (impulses) arise at some complex function of the amplitude and phase of the slow, graded synaptic potentials, their electrotonic spread to the spike-initiating locus and the curve of recovery of that locus. Besides the summation of synaptic events of different size, polarity, and frequency dependence, integration in these cells is shown by different kinds of afterdischarge and rebound following cessation of input.

Cells which are setting a pace and are not being synaptically driven show the same type of gradual, depolarizing, pacemaker potential between discharges as the pacemaker regions of the vertebrate heart. At least in some, the locus of this process is outside the cell body and there may be more than one rhythm in the same neuron. Although there is, apparently, no spike feedback, followers exert a positive feedback effect upon the pacemakers through some specific electric connection capable of spreading slow potentials electrotonically. Currents imposed either through such connections or externally accelerate the rhythm in one polarity and inhibit it in the other.

The effects of the inhibitor and accelerator nerve fibers are not mirror images; the former adapts considerably, has a brief aftereffect and rebound; the latter adapts little, has a long aftereffect and no rebound. Inhibition cannot be countered by acceleration. Each affects various aspects of the activity of the ganglion in different proportions, for example, burst frequency, spike frequency within the burst, and spike numb

ac

ac

su

region which may have some role in regulating heart beat. *See* NERVOUS SYSTEM (INVERTEBRATE). [T.H.B.]

Bibliography: T. H. Bullock, Parameters of integrative action of the nervous system at the neuronal level, *Exp. Cell Research*, suppl., 5:323-337, 1958.

Heart disorders

Any pathology of the heart, an organ located in the chest between the lungs and above the diaphragm. The heart is a hollow, muscular structure divided into right and left halves by a vertical partition of muscle. These halves are subdivided, by an incomplete transverse partition, into four chambers. The two larger thick-walled chambers below are called the ventricles. The two smaller thin-walled chambers formed above are the auricles. The three layers of the heart that may react to injury are the epicardium, the outer layer of the heart formed by the inner layer of the pericardium; the myocardium, the middle and muscular layer; and the endocardium, the thin innermost layer. The heart contains a supportive network of connective tissue in which blood vessels, lymphatics, and nerve fibers course.

Carditis. Any inflammation involving the heart is known as carditis. Rheumatic heart disease, a common and serious form of carditis, is a generalized systemic disease involving the collagen structures of the body. The reaction is considered to be an allergic response of the tissues previously sensitized by certain streptococci infections. In this hypersensitive state, the most serious lesions occur within the cardiovascular system. The degree of injury to the structures of the heart varies directly with the severity and number of streptococcal infections. The disease occurs most frequently in children and young adults; first attacks occur beyond the fourth decade of life infrequently. Although all three layers of the heart are involved in acute rheumatic carditis, the supportive system is most seriously affected. When present, the characteristic lesion seen microscopically is the rheumatic nodule or Aschoff body, the formation of which is one of the manifestations of degenerative changes that occur in the collagen structures of the supportive system. The collagen fibers undergo swelling and edema accompanied by an inflammatory infiltrate (plasma cells, lymphocytes, and occasional polymorphonuclear leukocytes). As the rheumatic activity subsides, the inflammatory elements are replaced by a proliferation of young fibroblasts. These changes are particularly evident around blood vessels within the myocardium. The ... is scar formation

ologi

ctive

TISSUE; EDEMA; HYPERSENSITIVITY; RHEUMATIC FEVER.

ENDOCARDIUM

The endocardium is a smooth, thin membrane lining the chambers and the four valves of the heart. On the right side, the tricuspid valve guards the common opening between the right auricle and right ventricle. It is formed by three strong, triangular folds of connective tissue covered by endocardium. On the left side, there is the mitral (or bi-

cuspid) valve formed by two such folds guarding the opening between the left auricle and left ventricle. Attached to the undersurfaces of both of these valves are stringlike structures called chordae tendineae. These chordae tendineae are connected by nipplelike projections (papillary muscles) to the thick muscle mass of the ventricles. The chordae tendineae and papillary muscle structures act as anchoring devices, preventing eversion of the valve leaflets during ventricular contraction. At the point where the pulmonary artery and aorta arise from their respective ventricles, three pouchlike folds form the pulmonary semilunar and aortic semilunar valves.

The supportive system of the heart acts as a flexible binding between the structural units of this organ. Near the valves of the heart, this binding forms dense circular bands of connective tissue which prevent excessive stretching of the valve openings when the chambers contract. These ringed bands also provide anchoring points for the myocardial fibers.

Endocarditis. This term is used to denote changes that occur in the heart valves affected by rheumatic disease. The basic pathological process is a valvulitis. Rheumatic valvulitis is the most common cause of deformed valves. In acute rheumatic endocarditis, small wartlike nodules called verrucae may appear along the line of closure of the valve leaflets. They occur on the surface of the valve leaflets that oppose the direction of the flow of blood through the heart, that is, on the auricular surface of the mitral and tricuspid valves and on the ventricular surface of the semilunar valves. Early in their formation the verrucae are small and translucent; later many capillaries and fibroblasts grow into them, securing them to the damaged valve leaflet. Recurrent episodes of rheumatic activity lead to the formation of new verrucae which may extend to the endocardium adjacent to the valve as well as the chordae tendineae. As new connective tissue is laid down in the process of repair, the involved valve leaflets and chordate tendineae become thickened and deformed. If deformity is severe, impairment of their function is assured. There are two types of deformity that may occur in chronic endocarditis, fusion of the valve leaflets and consequent narrowing (stenosis) of the valve orifice, and imperfect closure of the valve orifice (valvular insufficiency).

In stenosis, the affected chamber is unable to empty its blood completely in systole. This results in overfilling of the chamber in the next phase, diastole, of the cardiac cycle, when it receives its usual volume of blood. The compensatory mechanisms of dilatation and hypertrophy are set into motion in order to overcome the impedance of flow through the narrowed valve orifices.

Valvular insufficiency is caused by a failure of the scarred, contracted valve leaflets to coaptate properly, thus allowing for some degree of reverse flow of the blood or regurgitation. This occurs in the diastolic phase of the chamber affected when some of the blood ejected in the preceding systole leaks backward through the damaged valve and mixes with the new volume of blood received by the chamber. The overfilled chamber is distended by the increased volume of blood and thus the mechanisms of dilatation and hypertrophy are initiated. Adequate circulation of the blood by the heart can be maintained through these compensatory mechanisms as long as the increased metabolic requirements of the overworked heart muscle are met by increased coronary blood flow. Cardiac decompensation exists when these mechanisms can no longer compensate for the functional derangements imposed upon the heart by disease processes. Some of the clinical manifestations of cardiac decompensation are shortness of breath (dyspnea), swelling of the tissue (edema), accumulation of fluid in serous cavities of the body, enlargement of the heart, and limitation of physical activity.

Infectious endocarditis. Infectious endocarditis or bacterial endocarditis is the term applied to all types of bacterial infections involving the valves of the heart. It may be either acute or subacute in form. In the acute type, the patient usually succumbs to the disease in 6–8 weeks. In the subacute form, the patient may go beyond the 6–8-week period and even months may elapse before the clinical manifestations of bacterial endocarditis are evident. Acute bacterial endocarditis (also called ulcerative, malignant, and infectious endocarditis) is caused by the more virulent types of microorganism as opposed to the subacute form. The causative agents may be hemolytic streptococci, *Staphylococcus aureus*, *Pneumococcus*, gonococci, and colon bacilli. The microorganisms can usually be recovered in smears taken at postmortem examination from the vegetations on the valves. These vegetations are usually found on valves which have been previously damaged by rheumatic valvular disease or show congenital abnormalities. However, infrequently, they may also involve a normal valve. Sometimes the valvular vegetations are small and are similar in appearance to those of rheumatic endocarditis. Generally, they are large, gray, and friable. Ulceration is frequent and perforation of the leaflets of the valve may also occur. The vegetation is composed of red blood cells, fibrin, platelets, white blood cells, and entrapped colonies of bacteria. Dead tissue and ulceration is frequently found at the base of the lesion.

Subacute bacterial endocarditis. Subacute bacterial endocarditis, also known as endocarditis lenta, is more frequent than the acute form and the causative agent, in almost all cases, is *Streptococcus viridans*, an organism of relatively low virulence. Other causative organisms are meningococcus, brucella, influenza, and anthrax bacilli. In the subacute type, previous valvular damage is almost always evident and not infrequently coincidental active rheumatic disease is present. The affected heart valves show large, friable vegetations of reddish-gray color which may extend to the adjacent endocardium. Involvement of the myocardium by very minute abscesses may also be found. The

vegetation consists primarily of platelet masses near the surface of the valve and bacterial colonies with multinucleated and mononuclear cells near the base. In repair, scar tissue in varying stages of organization and sometimes calcification may be found within the vegetation. The clinical complications of bacterial endocarditis are toxemia, embolization by detached fragments of the vegetation to [illegible] kidney), and [illegible] eath was the [illegible] s of recovery were rare. With the widespread use of antibiotics, however, the incidence of disease has been diminished. *See* ANTIBIOTIC; EMBOLUS.

Syphilitis valvulitis. This disease is caused by the spirochete *Treponema pallidum*, which characteristically involves the aortic valve and wall of the aorta. In the tertiary stage, it is a granulomatous type of inflammation that causes scarring and retraction of the valve leaflets and dilatation of the aortic ring to which the valves are attached. Valvular insufficiency may occur because of aneurysmal dilatation of the aorta caused by the destruction of the elastic fibers contained within the wall of the aorta. Death may be due to fatal hemorrhage from rupture of the aortic aneurysm or cardiac failure secondary to aortic insufficiency. *See* HEMORRHAGE; SYPHILIS.

MYOCARDIUM

The myocardium, the muscle layer of the heart, by fusion of successive fibers, forms a sheath uniting each pair of auricles and ventricles. This layer supplies the force required to pump the blood through the miles and miles of blood vessels.

Any inflammatory condition involving the myocardium is myocarditis. Myocarditis is common in the acute phase of rheumatic fever in which Aschoff bodies and inflammatory infiltrates within the myocardium and adjacent to the walls of blood vessels are seen.

Infectious myocarditis. Inflammation of the cardiac muscle may arise from local or generalized infectious disease. Acute upper respiratory infection may give rise to myocardial damage which clinically may be so subtle that it can be recognized only by changes in the electrocardiogram. Streptococcal infection, gonococci, meningococci, and pneumococci may also cause inflammatory lesions of the myocardium. Between the myocardial fibers and the connective tissue surrounding the blood vessels of the heart there are inflammatory infiltrates composed of lymphocytes, polymorphonuclear leukocytes, plasma cells, and at times, large numbers of eosinophiles. In some cases, small microscopic abscesses are present within the area of the inflammatory infiltrates. Tuberculous pericarditis may sometimes spread to involve the myocardium. Parasitic invasion of the heart muscle, for example, trichinosis, may be another cause of myocarditis. Myocarditis may also be seen in certain viral diseases such as poliomyelitis, mumps, and measles. There is an idiopathic form of myocarditis, Fiedler's myocarditis, in which acute progressive myocardial failure occurs in an individual who had no evident primary disease. Many poisons affect the heart as part of their general effect on the body. The resulting condition is chemical myocarditis. *See* MEASLES; MUMPS; POLIOMYELITIS; TRICHINOSIS.

Myelosclerosis. A form of myocardial scarring that is preceded and caused by degenerative changes in the myocardial fibers is myelosclerosis. The cellular infiltrate that follows this myocardial change is of a lesser magnitude than that seen in the preceding types of inflammatory lesion. The muscle fibers take on a glassy, transparent appearance and fat becomes visible within the myoplasm. The inflammatory cellular infiltrates are secondary to these changes. The destroyed muscle is replaced by scar tissue which is seen around blood vessels and between muscle fiber structures in its distribution. Myocardial degeneration of this type is seen in diphtheritic myocarditis and in typhoid fever. These changes are thought to represent the toxic effects of the products of metabolism of the bacterial agents on the heart muscle. The complications of this type of myocarditis are the formation of blood clots on the damaged wall of the heart and the formation of emboli from these sites. *See* TYPHOID FEVER.

Myocardosis. Myocardosis includes the noninflammatory disorders that cause scarring of the heart muscle. The basic pathological process is usually arteriosclerosis of the coronary arteries. In this slow progressive vascular disorder, the blood supply to the muscle fibers is impaired. The affected muscle fibers waste away and are replaced by scar tissue. If the larger branches of the coronary system are affected, the degree of scarring is extensive and a large portion of heart wall may be involved including the endocardium. In acute coronary thrombosis, the involved myocardium takes on a reddish-brown to yellow appearance because of sudden obstruction of coronary blood flow to the area and consequent death (infarction) of large amounts of muscle tissue. The rapidity of healing is dependent upon the size of the infarct. The dead muscle tissue is removed by scavenger cells which invade the area, and about the third or fourth week, connective tissue proliferation into the area is evident. In about 2 months, repair is complete and scar formation is the end result. Another cause of myocardial scarring is coronary occlusion due to embolism from detached vegetations on the aortic or mitral valves as is sometimes seen as a complication in acute bacterial endocarditis. Myocardial scarring is also seen in hypertensive cardiovascular disease because coronary artery sclerosis is a common complication of this disorder. *See* INFARCTION; THROMBOSIS.

Angina pectoris. Angina pectoris is a clinical entity characterized by an oppressive or crushing type of chest pain originating beneath the breast bone. It is an expression of insufficient blood supply to the heart muscle due either to coronary

sclerosis or to spasm of the coronary arteries. The pain may be transmitted to the left shoulder and extend down the left arm. The attack is usually sudden in onset and may be of several minutes' duration. Severe apprehension and a feeling of impending disaster is usually associated with the attack. Relief is often obtained by cessation of activity or the taking of appropriate medication (coronary dilators). Anginal attacks may be precipitated by sudden changes in temperature, physical exertion, and emotional upset. The pain impulses are transmitted by the cardiac plexes of nerves.

PERICARDIUM

The pericardium is an inelastic, double-layered, saclike structure surrounding the heart. The inner layer of the pericardium is fused to the surface of the heart, forming the epicardium. Between the inner and the outer layers of the pericardium, there is a potential space called the pericardial cavity that normally contains 5–10 ml of clear fluid. This fluid acts as a lubricant between the two layers, allowing them to move smoothly upon each other with each movement of the heart. The pericardium sometimes serves to protect the heart from adjacent diseased structures in the chest

Pericarditis. Any inflammation of the pericardium which may be acute or chronic in type is called pericarditis. It is usually secondary to some other disease process within the body. It is frequently a complication of rheumatic carditis or of infections of adjacent structures in the chest, such as viral or bacterial pneumonias. It may also be seen as a terminal manifestation in chronic disease involving other organs, such as uremia of renal failure. Less frequently, pericarditis may be caused by secondarily infected penetrating wounds of the chest or by a malignant tumor involving the pericardium. The pericarditis may be one of several types. The common ones are fibrinous pericarditis, serofibrinous pericarditis, purulent pericarditis, hemorrhagic pericarditis, and chronic pericarditis.

Fibrinous pericarditis. The most common type of pericarditis, seen in rheumatic carditis and acute infectious diseases, is fibrinous pericarditis. Abundant fibrin and white blood cells are deposited between the layers of the pericardium. This imparts a yellow shaggy appearance to the surface of the heart and the end result is the formation of adhesions between the layers of the pericardium and thickening of the pericardial layers.

Serofibrinous pericarditis. This usually follows the fibrinous form of pericarditis and, when accompanied by a disproportionate amount of serous fluid, is often referred to as pericarditis with effusion. The amount of fluid may vary from several hundred milliliters to several liters.

Purulent pericarditis. This form is caused by pus-producing bacteria. Here the white blood cells predominate over the fibrin content of the exudate.

Hemorrhagic pericarditis. In this inflammatory reaction, the red blood cells impart a hemorrhagic appearance to the exudate which also contains fibrin and white blood cells. The usual cause of this type of pericarditis is either bacterial infection of the type that breaks down (hemolyzes) red blood cells or involvement of the pericardium by malignant tumor. Tuberculosis of the pericardium, an extension of tuberculosis of contiguous structures in the chest such as the lungs, lymph nodes, or bones may also cause a hemorrhagic type of pericarditis

Chronic pericarditis. A sequela of acute pericarditis is chronic pericarditis, in which there is an extensive degree of pericardial scarring. Serious circulatory disturbances may be a complication of chronic pericarditis if the veins which leave and enter the heart are compressed by contracting bands of scar tissue. Compression of these large thin-walled venous channels results in inadequate filling of the heart and, consequently, diminished output (constrictive pericarditis). This in turn results in an elevation of the blood pressure at the venous end of the capillaries of the greater and lesser circulation with the escape of excess fluid into the tissue spaces and serous cavities of the body Embarrassment of cardiac function is also seen in pericarditis with effusion The sudden and excessive accumulation of fluid in the inelastic pericardial sac exerts a constricting pressure on the heart (cardiac tamponade), and if not relieved, results in circulatory failure, shock, and death. Dissecting aneurysms of the aorta that extend into the pericardium and rupture of the heart wall following acute coronary occlusion may result in acute cardiac tamponade and sudden death. See Death

Dilatation and hypertrophy. Disease processes, directly or indirectly affecting any of the components of the cardiovascular system, may cause serious mechanical problems. This is particularly true if the process alters the balance between the volumes of blood delivered to and from the heart. Through the mechanisms of dilatation and hypertrophy the heart is able to compensate to a degree for limitation imposed on its function. Any disease process that results in increased filling of any of the cardiac chambers has a stretching effect on the muscle fibers of the involved chamber. The physiological result of stretching of the heart muscle is a more powerful contraction of the chamber wall in systole. This stretching effect (dilatation) continues until the cardiac chamber is able to accommodate and discharge the increased volume of blood delivered into it. In strenuous physical activity, dilatation may be temporary and reversible (physiologic dilatation). If the causes that initiate the mechanisms of dilatation persist for a considerable period of time (pathological dilatation), the muscle fibers undergo an increase in size through an increase in their cytoplasmic content. This change, termed hypertrophy, also augments the contractile power of the myocardium. Hypertrophy, in time, may be of such magnitude as to result in doubling, and in extreme cases, even tripling of the weight

the heart. These changes are particularly striking in the ventricles which reflect the amount of work demanded of them. See CARDIOVASCULAR SYSTEM; CIRCULATION DISORDERS; HYPERTROPHY. [M.A.C.]

Heartwater disease

A septicemic, infectious disease of cattle, sheep, and goats in equatorial and southern Africa caused by *Coudria ruminantium*, a rickettsialike pathogen. It is carried chiefly by the bont tick, *Amblyomma heberaeum* (see ACARINA). Man is not susceptible. See RICKETTSIALES.

Nonapparent infections in certain wild antelope, in addition to range stock, are thought to provide a reservoir in bushveld areas for new generations of young ticks since infection does not pass through tick eggs. Strains of varying virulence cause subacute or acute disease with high mortality. The name is derived from the common lesion of the heart, hydropericarditis. Clumps of organisms are seen in Giemsa-stained cells lining the blood vessels. Young animals are immunized by intentional infection, followed by treatment with sulfonamide or antibiotic drugs. [C.B.P.]

Heat

For the purposes of thermodynamics, it is convenient to define *all* energy while in transit, but unassociated with matter, as either heat or work. Heat is that form of energy in transit due to a temperature difference between the source from which the energy is coming and the sink toward which the energy is going. The energy is not called heat *before it starts to flow nor after it has ceased to flow* (see INTERNAL ENERGY). A hot object does contain energy, but calling this energy heat as it resides in the hot object can lead to widespread confusion.

Heat is a result of a potential difference between the source and sink called temperature. See BRITISH THERMAL UNIT (BTU); CALORIE. Work is energy in transit as a result of a difference in any other potential such as height. Work may be thought of as that which can be completely used for lifting weights. Heat differs from work, the other type of energy in transit, in that its conversion to work is limited by the fundamental second law of thermodynamics, or Carnot efficiency. This natural law is that the fraction of the heat Q convertible to work is determined by the relation $dW = Q(dT/T)$ for processes where the source and sink are but differentially different in temperature, or by the relation $dW = dQ(T_1 - T_2)/T_1$ where the source (at T_1) and the sink (at T_2) differ by a finite temperature interval.

For the above relations to be valid, temperature must be expressed on a thermodynamic temperature scale. Conversely, any temperature scale for which the above relations are valid, irrespective of the substance or material under investigation, is a thermodynamic [illegible] kine equals degrees Fahrenheit plus 456 67°) are thermodynamic temperature scales to the extent to which the constants 273.15° and 456 67° are known. The true nature of heat continues to be the subject of widespread speculation from a philosophical viewpoint. [H.C.W.]

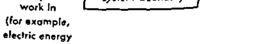
Bibliography: P. W. Bridgman, *The Nature of Thermodynamics*, 1941; H. C. Weber and H. P. Meissner, *Thermodynamics for Chemical Engineers*, 2d ed., 1957.

Heat balance

A special form of the total energy balance based on the principle of the conservation of energy. One statement about the principle of conservation of energy is, "energy can neither be created nor destroyed." To employ an energy balance in the analysis of a system, all of the energy terms involved must be taken into account. The different forms of energy entering a system are equated to the different types of energy leaving the system, plus or minus any change in the stored energy within the system. The illustration shows this heat balance.

In many processes the most significant energy terms are work and heat transfer. For systems of this type, the principle of the conservation of energy is known also as the first law of thermodynamics. The first law may be expressed in general terms as: (all heat transferred to the system and all work performed on the system) + (all energy associated with material flowing into the system) = (all heat transferred from the system and all work performed by the system) + (all energy associated with material flowing out of the system) ± (change in energy internal to the system).

In many steady continuous processes the change in the energy internal to the system is negligible. The work terms are only those involved as a result of forcing the material into and out of the process. Under these conditions the first law equation becomes one in which the heat transferred is equal to the difference in enthalpy between the material entering and leaving the system. This equa-



A thermodynamic system, such as a domestic heating system, illustrates heat balance. All material, heat, and work that enter the system either remain in it or leave it.

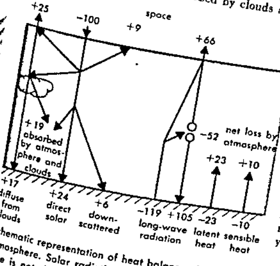
tion is often referred to as a heat balance. The heat balance is one of the most common applications of the energy balance. *See* ENTHALPY. [G.A.H.]

Heat balance, terrestrial atmospheric

The disposition, distribution, and transformation of the heat received by the earth-atmosphere system from the sun. A square centimeter area at the earth's mean distance from the sun and perpendicular to the sun's rays would receive about 2 calories (cal) of solar radiation per minute if there were no intervening atmosphere. This quantity is called the "solar constant." Since the area of a sphere is four times that of the circle it presents to parallel radiation, the outside of the earth's atmosphere receives an average of ½ cal/(cm²)(min) or ½ langley/min. One-third of this radiation is reflected and scattered back to space, mostly by clouds (*see* INSOLATION; METEROLOGICAL OPTICS). The remaining two-thirds is absorbed by the earth, clouds, and atmosphere and acts to raise their temperature and to evaporate water from the oceans and clouds.

The earth and atmosphere can lose heat to space only by radiation. Since the temperature changes on the earth over long periods of time are small compared to the changes that would have occurred had a significant fraction of the solar radiation been retained as a net heat gain, the outward radiation must be equal to the two-thirds of the solar radiation that is not reflected or scattered directly to space. *See* RADIATION, TERRESTRIAL.

H. G. Houghton's estimate of the disposition of the one-half langley per minute of solar radiation, taken as 100 units, is shown by the arrows on the left-hand side of the illustration. Of the 100 units, about 25 are reflected and scattered by clouds and about 9 units by the earth's surface and atmosphere. About 19 units are absorbed by clouds and



Schematic representation of heat balance of earth and atmosphere. Solar radiation reflected from earth's surface is not shown separately, but is included in solar radiation returned to space. (After H. G. Houghton)

atmosphere, and the remainder are absorbed at the surface of the earth. The middle part of the illustration shows the fluxes of long-wave terrestrial radiations at the top and bottom of the atmosphere. The atmosphere loses about 52 more units to earth and space than it gains from the earth (*see* GREENHOUSE EFFECT, TERRESTRIAL). The earth also transfers heat to the atmosphere by evaporation of water from the earth's surface followed by condensation in the atmosphere (release of latent heat) and by convection from the heated ground surface (sensible heat).

There are, of course, marked geographical deviations from this average picture of the heat balance. In particular, the tropical regions of the earth receive considerably more solar radiation than is lost by terrestrial radiation to space, and the polar regions considerably less. To maintain balance, heat must be transferred from low to high latitudes by the wind systems and, to some extent, by the ocean circulations. [L.D.K.]

Bibliography: H. G. Houghton, On the annual heat balance of the Northern Hemisphere, *J. Meteorol.*, 11(1):1–9, 1954.

Heat capacity

The quantity of heat required to raise a unit mass of homogeneous material one unit in temperature along a specified path, provided that during the process no phase or chemical changes occur. The unit mass may be 1 gram, 1 pound, or 1 gram-molecular weight (1 mole). Moreover, the path is so restricted that the only work effects are those necessarily done on the surroundings to cause the change to conform to the specified path. The path, except as noted later, is at either constant pressure or constant volume. This definition conforms to an average heat capacity for the chosen unit change in temperature.

Instantaneous heat capacity at a particular temperature is defined as the rate of heat addition relative to the temperature change at the temperature in question; that is, on a plot of heat addition Q as a function of temperature T, instantaneous heat capacity is given by the slope of the curve at the temperature in question. Units of heat capacity are energy units per unit mass of material per unit change in temperature.

In accordance with the first law of thermodynamics, heat capacity at constant pressure C_p is equal to the rate of change of enthalpy with temperature at constant pressure $(\partial H/\partial T)_p$ (*see* ENTHALPY). Heat capacity at constant volume C_v is the rate of change of internal energy with temperature at constant volume $(\partial U/\partial T)_v$ (*see* INTERNAL ENERGY). Moreover, for any material, the first law yields the relation

$$C_p - C_v = \left[P + \left(\frac{\partial U}{\partial V}\right)\right]_T \left(\frac{\partial U}{\partial T}\right)_P$$

Gases. For one mole of a perfect gas, the preceding relation becomes $MC_p - MC_v = R$, where M is the molecular weight of the gas under disc

and R is the perfect gas law constant. For gases, the ratio C_p/C_v is usually designated by the symbol K.

For monatomic gases at moderate pressures, MC_v is about 3, K is about 1.67, and the heat capacity changes but little with temperature. For diatomic gases, MC_v is approximately 5 at 20°C and moderate pressures. Change of heat capacity with temperature is usually small. The value of K is between 1.40 and 1.42. For triatomic gases at moderate pressures, MC_v varies from 6 to 7 and changes rapidly with temperature. The value of K varies, but is always smaller than that for the less complex molecules at the same conditions of pressure and temperature.

For gases with more than three atoms per molecule, no generalizations are reliable. However, as molecular complexity increases, heat capacity increases, the influence of temperature on heat capacity increases, and K decreases. Figure 1 shows average MC_p for several common gases. Up to pressures of a few atmospheres, the effect of pressure on heat capacity of gases is small and is usually neglected.

Solids. For solids, the atomic heat capacity (heat capacity when the unit mass under discussion is 1

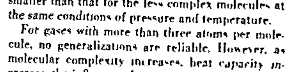


Fig. 1 Variation of average molar heat capacity with temperature for gases (Based on data of H. C. Hottel, MIT)

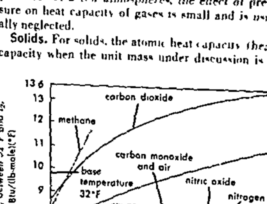


Fig. 2. Atomic heat capacity as a function of temperature. (Based on data of G. N. Lewis and G. E. Gibson, J. Am. Chem. Soc., 39:2554–81, 1917)

atomic weight) may be closely approximated by an equation of the type

$$C_v = f\left(\frac{T}{\theta}\right)^n$$

where $n = 1$ for elements of simple crystalline form, but has a smaller value for those of more complex structure; θ is characteristic of each element; f is a function that is the same for all substances; and T is the absolute temperature. Figure 2 compares measured values with calculated values.

For all solid elements at room temperature, C_v is about 6.4 calories per gram atom per °C. This approximation may be used when no experimental data are available, but errors may be considerable, particularly for elements with atomic weights less than 39. Kopp's law states that for solids, the molal heat capacity of a compound at room temperature and pressure approximately equals sum of heat capacities of the elements in the compound. Errors are considerable, but may be reduced by judicious choice of atomic heat capacities for the lighter elements. Recommended values for some of these are given in the following list of constants for Kopp's law:

Element	*Heat capacity*
All heavy elements	6.4
Boron	2.7
Carbon	1.8
Fluorine	5.0
Hydrogen	2.3
Oxygen	4.0
Phosphorus and sulfur	5.4
Silicon	3.5

Use of Kopp's law is justified only when no experimental data are available.

Figures 3 and 4 give instantaneous heat capacities for some industrially important solids.

Liquids. For liquids and solutions, no useful generally applicable approximations are available. For aqueous solutions of inorganic salts, the approximate heat capacity of the solution may be estimated by assuming the dissolved salt to have negligible heat capacity. Thus, in a 20% by weight solution of any salt in water 0.8 would be the estimated heat capacity.

Effect of pressure on heat capacities at any temperature may be calculated by the relations

$$\left(\frac{\partial MC_p}{\partial P}\right)_T = -T\left(\frac{\partial^2 V}{\partial T^2}\right)_P$$

and

$$\left(\frac{\partial MC_v}{\partial V}\right)_T = T\left(\frac{\partial^2 P}{\partial T^2}\right)_V$$

Heat capacities at constant temperature. Not so familiar as C_p and C_v is the heat necessary to cause unit change in pressure in a unit mass of material at constant temperature, and the heat required ... constant tempera- ... /∂V. heat

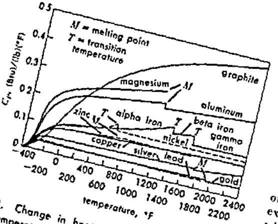


Fig. 3. Change in heat capacity of some elements with temperature. (Based on data of K. K. Kelly, U.S. Bur. Mines Bull. 371, 1934)

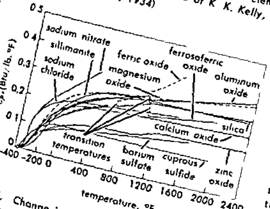


Fig. 4. Change in heat capacity of compounds with temperature. (Based on data of K. K. Kelly, U.S Bur. Mines Bull. 371, 1934)

capacities. *See* SPECIFIC HEAT; THERMODYNAMIC PRINCIPLES. [H.C.W.]

Bibliography: H. C. Weber and H. P. Meissner, *Thermodynamics for Chemical Engineers*, 2d ed., 1957.

Heat exchanger

A device used to transfer heat from a fluid flowing on one side of a barrier to another fluid (or fluids) flowing on the other side of the barrier.

When used to accomplish simultaneous heat transfer and mass transfer, heat exchangers become special equipment types, often known by other names. When fired directly by a combustion process, they become furnaces, boilers, heaters, tube still heaters, and engines. If there is a change in phase in one of the flowing fluids—condensation of steam to water, for example—the equipment may be called a chiller, evaporator, sublimator, distillation-column reboiler, still, condenser, or cooler condenser.

Heat exchangers may be so designed that chemical reactions or energy-generation processes can be carried out within them. The exchanger then becomes an integral part of the reaction system and may be known, for example, as a nuclear reactor, catalytic reactor, or polymerizer.

Heat exchangers are normally used only for the transfer and useful elimination or recovery of heat without an accompanying phase change. The fluids on either side of the barrier are usually liquids, but they may also be gases such as steam, air, or hydrocarbon vapors; or they may be liquid metals such as sodium or mercury. Fused salts are also used as heat-exchanger fluids in some applications.

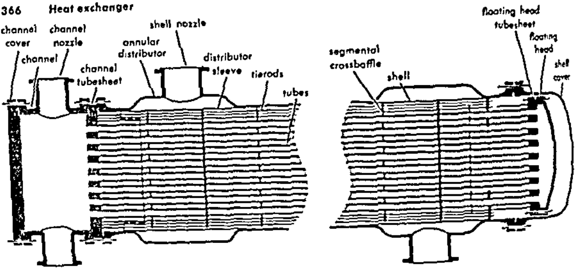
With the development and commercial adoption of large, air-cooled heat exchangers, the simplest example of a heat exchanger would now be a tube within which a hot fluid flows and outside of which air is made to flow for the purpose of cooling. By similar reasoning, it might be argued that any container of a fluid immersed in any fluid could serve as a heat exchanger if the flow paths were properly connected, or that any container of a fluid exposed to air becomes a heat exchanger when a temperature differential exists. However, engineers will insist that the true heat exchanger serve some useful purpose that the heat recovery be meaningful or profitable.

Most often the barrier between the fluids is a metal wall such as that of a tube or pipe. However, it can be fabricated from flat metal plate or from graphite, plastic, or other corrosion-resistant materials of construction. If, as is often the case, the barrier wall is that of a seamless or welded tube, several tubes may be tied together into a tube bundle (see diagram) through which one of the fluids flows distributed within the tubes. The other fluid (or fluids) is directed in its flow in the space outside the tubes through various arrangements of passes. This fluid is contained by the heat-exchanger shell. Discharge from the tube bundle is to the head (heads) and channel of the exchanger. Separation of tube-side and shell-side fluids is accomplished by using a tube sheet (tube sheets).

Applications. Heat exchangers find wide application in the chemical process industries, including petroleum refining and petrochemical processing; in the food industry, for example, for pasteurization of milk and canning of processed foods; in the generation of steam for production of power and electricity; in nuclear reaction systems; in aircraft and space vehicles, and in the field of cryogenics for the low-temperature separation of gases. Heat exchangers are the workhorses of the entire field of heating, ventilating, air-conditioning, and refrigeration.

Classifications. The exchanger type described in general terms above and illustrated by the diagram is the well-known shell-and-tube heat exchanger. Shell-and-tube exchangers are the most numerous, but constitute only one of many types. Today's exchangers range from the simple pipe within a pipe—with a few square feet of heat-transfer surface—up to the complex-surface exchangers that provide thousands of square feet of heat-transfer area.

In between these extremes is a broad field of shell-and-tube exchangers often specifically named by distinguishing design features; for example,



Schematic diagram of heat exchanger.

U tube, fin tube, fixed tube sheet, floating head, lantern-ring packed floating head, socket-and gland packed floating head, split-ring internal floating head, pull-through floating head, nonremovable bundle with floating head or U-tube construction, and bayonet type

Also, varying pass arrangements and baffle-and-shell alignments add to the multiplicity of available designs. Either the shell-side or tube-side fluids, or both, may be designed to pass through the exchanger several times in concurrent, countercurrent, or cross flow to the other fluid.

The concentric pipe within a pipe (double pipe) serves as a simple but efficient heat exchanger. One fluid flows inside the smaller-diameter pipe and the other flows, either concurrently or countercurrently, in the annular space between the two pipes, with the wall of the larger-diameter pipe serving as the shell of the exchanger.

To solve new processing problems and to find more economical ways of solving old ones, new types of heat exchangers are being developed. There has been much emphasis on cramming more heat-transfer surface into less and less volume. Extended-surface exchangers, such as those built with fin tubes, are finding wide application.

Water shortage has added a new dimension to heat-exchanger design. In its four newest refineries (all outside the United States) a large oil company has not used a single water-cooled exchanger. All are air-cooled. On the other hand, in the United States, two new plants on opposite banks of the same river disagreed on the merits of air-cooling. One chose water-cooling and the other air-cooling.

Plate-type heat exchangers, long used in the milk industry for pasteurization and skimming, are moving into the chemical and petroleum industries. Coiled tubular exchangers and coiled-plate heat exchangers are winning new assignments. Spiral exchangers offer short cylindrical shells with flat heads, carrying inlets and outlets leading to internal spiral passages. These passages may be made with spiral plates or with spiral banks of tubes. Exchangers with mechanically scraped surfaces are finding favor for use with very viscous and pastelike materials.

A somewhat unusual type of plate heat exchanger is one in which sheets of 16-gage metal, seam- and spot-welded together, are embossed to form transverse internal channels which carry the heat-transfer medium. This type is often used for immersion heating in electroplating and pickling.

Materials of construction. Every metal seems to be a possible candidate as a material of construction in fabrication of heat exchangers. Most often, carbon steels and alloy steels are used because of the strength they offer, especially when the exchanger is to be operated as a pressure vessel. Because of excellent heat conductance, brass and copper find wide use in exchanger manufacture.

Corrosion plays a key role in the selection of exchanger construction materials. Often, a high-priced material will be selected to contain a corrosive tube-side fluid, with a cheaper material being used on the less corrosive shell side.

For special corrosion problems, exchangers are built from graphite, ceramics, glass, bimetallic tubes, tantalum, aluminum, nickel, bronze, silver, and gold.

Problems of use. Each of the fluids and the barrier walls between them offers a resistance to heat transfer. However, another major resistance that must be considered in design is the formation of dirt and scale deposited on either side of the barrier wall. This resistance may become so great that the exchanger will have to be removed from service periodically for cleaning.

Chemical and mechanical methods may be used to remove the dirt and scale. For mechanical cleaning, the exchanger is removed from service and opened up. Perhaps the entire tube bundle is pulled from the exchanger shell if the plant layout has provided space for this to be done. If the deposit is on the inside of straight tubes, cleaning may be accomplished merely by forcing a long worm or wire brush through each tube.

More labor is required to remove deposits on the shell side. After removal of the tube bundle, special cleaning methods such as sandblasting may be necessary.

Much engineering effort has gone into the design of heat exchangers to allow for fouling. However, D. Q. Kern has suggested that methods are available to design heat exchangers that, by accommodating a certain amount of dirt in a thermal design, will allow heat exchangers to run forever without shutdown for cleaning. Commercial units designed in this fashion are operating today.

Another operating problem is allowance for differential thermal expansion of metallic parts. Most operating difficulties arise during the startup or shutdown of equipment. Therefore, M. S. Peters suggests the following general rules:

1. Startup. Always introduce the cooler fluid first. Add the hotter fluid slowly until the unit is up to operating conditions. Be sure the entire unit is filled with fluid and there are no pockets or trapped inert gases. Use a bleed valve to remove trapped gases.

2. Shutdown. Shut off the hot fluid first, but do not allow the unit to cool too rapidly. Drain any materials which might freeze of solidify as the exchanger cools

3. Steam condensate. Always drain any steam condensate from heat exchangers when starting up or shutting down. This reduces the possibility of water hammer caused by steam forcing the trapped water through the lines at high velocities.

Standardization. Users have requested that heat exchangers be made available at lower prices through the standardization of designs. Organizations active in this work are the Tubular Exchangers Manufacturers' Association, American Petroleum Institute, American Standards Association, and the American Institute of Chemical Engineers. *See* CONDENSER VAPOR; CONDUCTION (HEAT); CONVECTION (HEAT); COOLING TOWER; DISTILLATION; EVAPORATOR; FURNACE CONSTRUCTION; HEAT RADIATION; HEAT TRANSFER; TUBE-STILL HEATER.

[R.F.FR.]

Bibliography: W. E. Glausser and J. A. Cortright, How to specify heat exchangers, *Chem. Eng.*, 62(12):203–206, 1955; D. Q. Kern, Speculative process design, *Chem. Eng.*, 66(20):127–142, 1959; M. S. Peters, *Elementary Chemical Engineering*, 1954; J. C. Smith, Trends in heat exchangers, *Chem. Eng.*, 61(6):232–238, 1953.

Heat pump

The thermodynamic counterpart of the heat engine. A heat pump raises the temperature level of heat by means of work input. In its usual form a compressor takes refrigerant vapor from a low-pressure, low-temperature evaporator and delivers it at high pressure and temperature to a condenser (Fig. 1). The pump cycle is identical with the customary vapor-compression refrigeration system (*see* REFRIGERATION CYCLE).

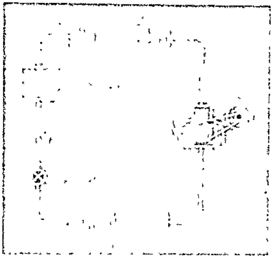


Fig. 1. Basic flow diagram of heat pump with motor-driven compressor For summer cooling, condenser is outdoors and evaporator indoors; for winter heating, condenser is indoors and evaporator outdoors.

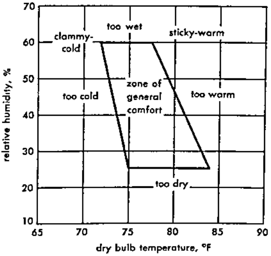


Fig. 2. Indoor climatic conditions acceptable to most people when doing desk work; continuous air motion with 5–8 air changes per hour.

Application to comfort control. For air-conditioning in the comfort heating and cooling of space, a heat pump uses the same equipment to

placing the low-temperature evaporator in the conditioned space during the summer and the high-temperature condenser in the same space during the winter (Fig. 3). Thus, if 70°F is to be maintained in the conditioned space regardless of the season, this would be the theoretic temperature of the evaporating coil in summer and of the condensing coil in winter. The actual temperatures on the

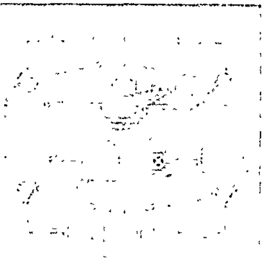


Fig 3. Schematic of air-to-air heat pump installation; fixed air circuit with valves in summer positions (dotted lines show winter positions).

refrigerant side of these coils would need to be below 70°F in summer and above 70°F in winter to permit the necessary transfer of heat through the coil surfaces.

If the average outside temperatures are 100°F in summer and 40°F in winter, the heat pump serves to raise or lower the temperature 30° and to deliver the heat or cold as required. The ultimate ideal cycle for estimating performance is the same Carnot cycle as that for heat engines. The coefficient of performance cp_c as a cooling machine is

$$cp_c = \frac{\text{refrigeration}}{\text{work}} = \frac{T_c}{T_h - T_c} \quad (1)$$

and the coefficient cp_w as a warming machine is

$$cp_w = \frac{\text{heat delivered}}{\text{work}} = \frac{T_h}{T_h - T_c} \quad (2)$$

where T is temperature in degrees absolute, and the subscripts c and h refer to the cold and hot temperatures respectively.

For the data cited, the coefficients of performance are

$$cp_c = \frac{460 + 70}{(460 + 100) - (460 + 70)} = 17.7$$

and $$cp_w = \frac{460 + 70}{(460 + 70) - (460 + 40)} = 17.7$$

The significance of these coefficients is that ideally, for 1 kilowatt-hour (kwhr) of electric energy input to the compressor there will be delivered $3413 \times 17.7 = 60{,}000$ Btu/hour as refrigeration or heating effect as required. This is a great improvement over the alternative use of resistance heating, typically, where 1 kwhr of electric energy would deliver only 3413 Btu. The heat pump uses the second law of thermodynamics to give a much more substantial return for each kwhr of electric energy input. The electric energy serves to move heat already present to a desired location.

Effect of seasonal loads. With the prevalent acceptance of summer comfort cooling of space, it is entirely possible and practical to use the same compressor equipment and coils for winter heating and for summer cooling and to dispense with the need for direct-fired apparatus using oil, gas, coal, or wood fuel.

For an economical installation, equipment must be of correct size for both the summer cooling and winter heating loads. Climatic conditions have a significant influence and can lead to imbalance on sizing. If the heating and cooling loads are equal the equipment can be selected with minimum investment. However, generally the loads are not balanced; in the temperate zone the heating load is usually greater than the cooling load. This necessitates (1) a large, high horsepower compressor fitted to the heating demand, (2) a supplementary heating system (electrical resistance or fuel), or (3) a heat-storage system.

If well water or the ground serves as the heat source, the imbalance is less severe than when atmospheric air is the source. However, the uncertain heat transfer rates with ground coils, and the impurities, quality, quantity, and disposal of water mitigate the use of these sources.

Atmospheric air as a heat source is preferable, particularly with smaller domestic units. A self-contained, packaged unit of this type offers maximum dependability and minimum total investment. The performance of such a unit for the heating cycle is illustrated in Fig. 4. Curves of heat required and heat available show the limitations on capacity. The heat delivered by the pump is less than the heat required at low temperatures, so that there is a deficiency of heat when the outside tem-

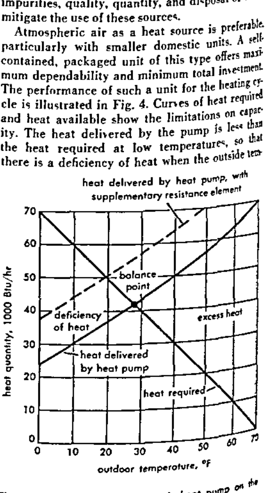


Fig. 4. Performance of air-to-air heat pump on the heating cycle.

perature goes below, in this case, about 28°F. The intersection of the two curves is the balance point. There is an area of excess heat to the right and of deficiency of heat to the left of the balance point. Many devices and methods are offered to correct this situation, such as storage systems, supplementary heaters, and compressors operating alternatively in series or in parallel.

In temperate regions heat-pump installations achieve coefficients of performance in the order of 3 on heating loads when all requirements for power, including auxiliary pumps, fans, resistance heaters, defrosters, and controls are taken into account. Automatic defrosting systems, when air is the heat source, are essential for best performance, with the defrost cycle occurring twice a day.

Heat pumps are uneconomical if used for the sole purpose of comfort heating. The direct firing of fuels is generally more attractive from an overall financial viewpoint. The investment in heat-pump equipment is higher than that for the conventional heating system. Unless the price of electric energy is sufficiently low, or the price of fuels very high, the heat pump cannot be justified solely as a heating device. However, if there is also need for comfort cooling of the same space in summer, the heat pump, to do both the cooling and heating, becomes attractive. This use of the heat pump will probably increase in the future with the broad acceptance of summer air conditioning represented by the sale of 1,000,000 units a year in the United States since 1950.

One can also expect that the heat pump will find increasing acceptance for operation in conjunction with solar heating systems. The efficiency of a solar collector rises rapidly with lower collection temperatures so that an electrically driven heat pump can use such a heat source with a high over-all coefficient of performance.

The heat pump is also used for a wide assortment of industrial and process applications such as low temperature heating, evaporation, concentration, and distillation. [T.B.]

Bibliography: S. J. Davies, *Heat Pumps and Thermal Compressors*, 1950; E. N. Kemler and S. Oglesby, *Heat Pump Applications*, 1950; P. Sporn, E. R. Ambrose, and T. Baumeister, *Heat Pumps*, 1947.

Heat radiation

The energy radiated by solids, liquids, and gases as a result of their temperature. Such radiant energy is in the form of electromagnetic waves and covers the entire electromagnetic spectrum, extending from the radio-wave portion of the spectrum, through the infrared, the visible, the ultraviolet, x-ray, and γ-ray portions. From most hot bodies on earth this radiant energy lies largely in the infrared region. See ELECTROMAGNETIC RADIATION; INFRARED RADIATION.

Radiation is one of the three basic methods of heat transfer, the other two methods being conduction and convection. See CONDUCTION (HEAT); CONVECTION (HEAT); HEAT TRANSFER.

A hot plate at 400°K (261°F) may show no visible glow; but a hand which is held over it senses the warming rays emitted by the plate. A temperature of more than 1000°K is required to produce a perceptible amount of visible light. At this temperature a hot plate glows red and the sensation of warmth increases considerably, demonstrating that the higher the temperature of the hot plate the greater the amount of radiated energy. Part of this energy is visible radiation and the amount of this visible radiation increases with increasing temperature. A steel furnace at 1800°K shows a strong yellow glow. If a tungsten wire (used as the filament in incandescent lamps) is raised by resistance heating to a temperature of 2800°K, it emits a bright white light. As the temperature of a substance increases, additional colors of the visible portion of the spectrum appear, the sequence being first red, then yellow, green, blue, and finally violet. The violet radiation is of shorter wavelength than the red radiation and it is also of higher quantum energy.

In order to produce strong violet radiation, a temperature of almost 3000°K is required. Ultraviolet radiation necessitates even higher temperatures and there is no solid on earth which can withstand such temperatures without melting. The sun emits considerable ultraviolet radiation, as evidenced by the sunburn it produces. The spectral distribution of the sun's radiation has been measured and the temperature of the sun's surface has been determined from Wien's displacement law and corresponds to about 6000°K (Wien's law is discussed later). Such temperatures have been produced on earth in gases ionized by electrical discharges. The mercury-vapor lamp used on highways, the fluorescent lamp used in offices, and the xenon compact-arc lamp used in searchlights are good examples of such gas discharges. They emit large amounts of ultraviolet radiation. Temperatures up to 20,000°K, however, are still much too low to produce x-rays or γ-radiation. Approaches to the utilization of nuclear energy have made use of the fusion of deuterons in magnetically constricted arcs at extremely high currents. By this means, temperatures above 1×10^6 °K have been obtained for small fractions of a second. These devices require enormous amounts of energy to produce such high temperatures. A gas maintained at such temperatures emits x-rays and γ-rays. See FUSION, NUCLEAR; SUN; ULTRAVIOLET RADIATION.

Theory. The emission of radiation is explained in terms of excited atoms and nuclei. For example, electrons in an atom can be ejected from their normal orbits around the atom into those farther from the nucleus. When this happens the atom is said to be in an excited state. This occurs when energy, supplied from outside a substance, is converted into thermal motion and finally into excitation. A short time after excitation, the electrons re-

turn to their normal orbits and give off their excess energy ΔE in the form of radiation of a particular frequency ν. This wavelength may be determined by the relation $\Delta E = h\nu$, where h is Planck's constant. *See* PLANCK'S CONSTANT; *see also* ATOMIC STRUCTURE AND SPECTRA.

In a gas, the thermal motion consists of substantially unhindered movement of the individual particles with different velocities. In a solid, on the other hand, the thermal motion is an oscillating movement of the particles, with varying displacements, about their fixed positions. The extent of the thermal motion depends upon the temperature. The hotter the substance, the greater the thermal motion and the higher the intensity and energy of the radiation. An energy distribution of the radiation intensity results, for example, from the distribution of velocities of the particles in a gas or from the distribution of displacements of the particles about their positions in a solid.

Further, the maximum available energy (excitation energy) depends upon temperature, and this explains why the energy of emitted radiation shifts to shorter wavelength (that is, higher energy) as the temperature is increased. For instance, a temperature of 1000°K produces just enough excitation energy for the dark red glow which contains the longest wavelengths within the visible portion of the spectrum. As explained before, higher temperatures or greater excitation energies are necessary to excite measurable quantities of the shorter wavelength regions. It is obvious that with decreasing temperatures, less excitation energy is available, and the amount of heat radiation decreases until finally at the absolute zero of temperature (0°K) substances radiate no energy because all atomic motion has ceased. However, for a definition of so-called zero point energy, *see* QUANTUM MECHANICS.

The radiated energy per second is commonly expressed in terms of joules per second, or watts. Other units often used are ergs per second or calories per second. These are related to each other as follows: 1 watt = 1 joule/sec = 10^7 erg/sec = 0.239 cal/sec. For instance, the sun radiates onto 1 cm² of the earth's surface 2 cal/min or 1/30 cal/sec or about 1/7 watt (see SOLAR CONSTANT). The total energy radiated from 1 cm² of a tungsten wire in an incandescent lamp at 2800°K is 112 watts. The same wire at room temperature emits only 0.0015 watt.

Energy distribution curves. In order to evaluate the usefulness of a heat radiator, energy distribution curves are used. These are graphs of relative or absolute radiated energy versus the wavelength of radiation (expressed in microns, 1×10^{-6} meter, or in angstroms, 1×10^{-10} meter) or the frequency (velocity of light/wavelength) expressed in cycles per second. Such graphs show how the energy radiated from a substance at a certain temperature is distributed over the various portions of the spectrum. The usefulness of these graphs lies in the fact that they provide information, for example, on

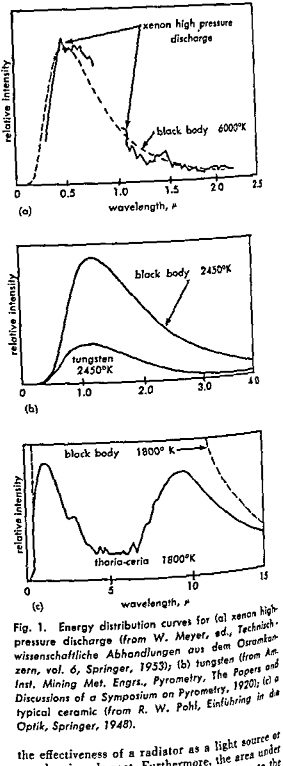


Fig. 1. Energy distribution curves for (a) xenon high-pressure discharge (*from W. Meyer, ed., Technisch-wissenschaftliche Abhandlungen aus dem Osramkonzern, vol. 6, Springer, 1953*); (b) tungsten (*from Am. Inst. Mining Met. Engrs., Pyrometry, The Papers and Discussions of a Symposium on Pyrometry, 1920*); (c) a typical ceramic (*from R. W. Pohl, Einführing in die Optik, Springer, 1948*).

the effectiveness of a radiator as a light source or as a heating element. Furthermore, the area under the energy distribution curve is equivalent to the total radiated energy. The energy distribution curves for tungsten metal, thoria plus 1% ceria (a ceramic), and the xenon high-pressure electrical gas discharge are illustrated in Fig. 1.

The energy distribution of various substances differs because of their internal properties and their surface condition. As a common rule, which holds well above 30,000 angstroms (A) or 3 microns (μ), substances with good electrical conduc-

tivity, especially metals, are poor emitters of radiation and are good reflectors (for example, silver or aluminum). Insulators radiate strongly in the infrared region of the spectrum and have gaps of low radiation intensity near the visible portion of the spectrum, as shown in Fig. 1c. These gaps are due to the electronic band structure of insulators (see INSULATOR, ELECTRIC). Roughening the surface of all radiators increases the emitted energy. This is true because tiny holes in the surface act as cavity radiators, radiating almost black-body energy. (Cavity radiators and black-body radiation are discussed later.)

Energy distribution curves are obtained by passing white (heterochromatic) light through a monochromator (quartz prism, grating, and the like) and measuring the spectral intensity of radiation with a phototube or a thermopile. The measured intensities at the various wavelengths are then plotted either as per cent of the maximum intensity (relative energy) or as absolute intensity (absolute energy) versus the wavelength. See RADIOMETRY.

A radiator used in heating rooms should produce much infrared radiation (heat) and no light, whereas much visible light and little heat is desired from a light source. Unfortunately, an energy distribution curve gives a true picture of the radiator for one particular temperature only. If more information is needed, a set of such curves would have to be provided for the temperature range of interest.

Black-body radiation. Because of the tedious experimental work involved in determining such curves, a different and more fruitful approach is generally taken. Two quantities characterize a heat radiator completely: the total emissivity and the spectral emissivity, which are designated by ϵ and ϵ_λ respectively, where the subscript λ designates wavelength. Both emissivities, in conjunction with the radiation properties of a black body, describe fully the behavior of a real heat radiator. The radiation properties of a black body are completely stated by Planck's radiation law.

Planck's law and the concept of black-body radiation are of utmost importance for the understanding of heat radiation. The black body signifies in the domain of heat radiation what any other standard, such as the standard meter, signifies in its own domain. A black body is defined as a body which emits the maximum amount of heat radiation. Although there exists no perfect black-body radiator in nature, it is possible to construct one on the principle of cavity radiation.

A cavity radiator is usually understood to be a heated enclosure with a small opening which allows some radiation to escape or enter. The escaping radiation from such a cavity has the same characteristics as black-body radiation.

As can be seen from Fig. 2, radiation energy which enters the cavity is almost completely absorbed because of the multiple reflections it encounters. This follows because at each reflecting point some of the energy is absorbed by the walls.

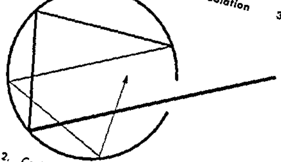


Fig. 2. Cavity radiator.

The absorptivity of the cavity hole is essentially unity, independent of the wall material. As a consequence of Kirchhoff's law, the emissivity of the cavity is also unity, and this fulfills the definition of a black-body radiator. In practice, the cavity is approximated by a small hole or even a wedge cut into a surface.

Kirchhoff's law. This law correlates mathematically the heat radiation properties of materials at thermal equilibrium. It is often called the second law of thermodynamics for radiating systems.

Kirchhoff's law can be expressed as follows: the ratio of the emissivity of a heat radiator to the absorptivity of the same radiator is a function of frequency and temperature alone. This function is the same for all bodies and it is equal to the emissivity of a black body. When ϵ is the emissivity of a real radiator, α its absorptivity, and $E = 1$ the emissivity of a black body, Kirchhoff's law takes the following form:

$$\epsilon/\alpha = E = 1$$

A substance, when brought without contact into an evacuated enclosure the walls of which are at a constant but higher temperature than the body, will assume the wall temperature after some time. However, it will not exceed it. Under these conditions, the exchange of energy can take place only by radiation. As the test body receives radiation from the walls it will absorb some of it, transforming it into motion of its elementary particles, and thereby raising its own temperature. Thermal equilibrium is obtained when the temperature of the walls and the test body is the same; in this case the test body must emit as much energy as it receives. If it absorbs all the impinging radiation, it is a black body. If it absorbs only a fraction of the impinging radiation, the other part must be reflected in order to maintain the equilibrium. These statements require that the absorptivity be equal to the emissivity. This is the form in which Kirchhoff's law is often stated. For opaque bodies, absorptivity plus reflectivity must be equal to unity, and therefore the emissivity and the absorptivity respectively must be unity minus the reflectivity. A consequence of Kirchhoff's law is the postulate that a black body has an emissivity which is greater than that of any other body.

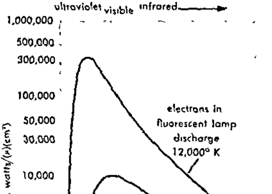


Fig. 3. Graphs of Planck's law for various temperatures.

Planck's radiation law. This celebrated law represents mathematically the energy distribution of the heat radiation from 1 cm^2 of surface area of a black body at any temperature. It is the only heat radiation law which is accurate throughout the entire spectrum. The basis of Planck's radiation law was experimental data obtained from measurements on cavity radiators.

Planck's radiation law has great importance. Formulated by Max Planck early in the twentieth century, it laid the foundation for the advance of modern physics and the advent of quantum theory. In determining the heat radiation of hot bodies Planck's radiation law is a basic tool in research and development, both in science and industry. The radiation law can be used to predict light output of incandescent lamps, the cooling time of molten steel, heat dissipation of nuclear reactors, the energy radiated from the sun, and the temperature of the stars; there are many other important applications.

Although Planck's law can be derived on theoretical grounds alone, it was deduced from experiment. Prior attempts to calculate the heat radiation of a black body had described the radiation as consisting of electromagnetic waves whose energy content could vary continuously. Those attempts did not match the experimental results. Planck replaced the concept of continuous energy with the idea that the energy existed in bundles; that is, the energy was quantized (see QUANTUM MECHANICS). This concept was a drastic innovation at that time. However, upon calculating the radiation, Planck found that the following expression described the experimental results completely

$$R_\lambda = 37{,}410/\lambda^5(e^{14{,}380/\lambda T} - 1)$$

This is the mathematical expression of Planck's radiation law, where R_λ is the total energy radiated from the body measured in watts per square centimeter per unit wavelength, at the wavelength λ. The wavelength in this formula is measured in microns. The quantity T is the temperature in °K and e is the base of the natural logarithms. Figure 3 presents graphs of Planck's law for various temperatures and shows substances which attain these temperatures. It should be noted that these substances will not radiate as predicted by Planck's law since they are not black bodies themselves.

As can be seen from Fig. 3, the radiation increases at every point of the energy spectrum as the temperature is increased. At all temperatures the energy radiated at the extremes of the energy spectrum approaches zero and has a maximum at some place in between. The total area under any of the curves, when plotted as in Fig. 1, measures the total energy radiated by the body at the temperature represented by the curve.

Three important aspects of Planck's radiation law can be examined. First, the behavior of the law at the extremes of the energy spectrum leads to a discussion of Wien's radiation law, the Rayleigh-Jeans radiation law, and the so-called ultraviolet [illegible] wavelength at

[illegible]

Wien's displacement law. Finally, the total amount of energy radiated at any temperature can be investigated. This leads to the Stefan-Boltzmann law. The four laws mentioned were well known prior to the formulation of Planck's law. It is the fact that the Planck law so neatly sums up the four earlier laws and introduces the implication of energy quantization which made it of such importance in the development of modern physics during the twentieth century.

Rayleigh-Jeans law. The heat radiation from a black body at long wavelengths is adequately described by the Rayleigh-Jeans radiation law. For large values of λT, Planck's law simplifies to the Rayleigh-Jeans law

$$R_\lambda = 2560\ T/\lambda^4$$

This law states that the energy radiated at any temperature increases without limit as the wavelength decreases. As can be seen from Figs. 1 and 3, this law can be accurate only for wavelengths much larger than that at which the maximum occurs. For wavelengths shorter than this maximum, the energy radiated from a black body actually decreases again. If a black body acted as predicted by the Rayleigh-Jeans law, then the energy radiated at very short wavelengths, in the ultraviolet region would become extremely large and the total energy radiated would be infinite. This is known as the ultraviolet catastrophe, and would be valid at any temperature, no matter how low.

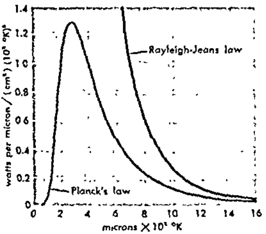


Fig. 4. Planck's law as expressed in the form of Wien's displacement law. The Rayleigh-Jeans law is shown for comparison.

Wien's radiation law. This law is valid at short wavelengths and is obtained from Planck's law by taking λT as very small. Planck's formula then becomes

$$R_\lambda = \frac{37{,}410}{\lambda^5} e^{-14{,}380/\lambda T}$$

This law is accurate in the visible region of the spectrum below 3000°K.

Wien's displacement law. This law is obtained from Planck's law by the process of differentiation. It describes the shift with temperature of the wavelength at which the maximum amount of radiation occurs by the following formula

$$\lambda_{max} T = 2898 \quad \text{(micron-degrees)}$$

Thus, the product of the temperature of a black body and the wavelength at which the maximum amount of radiation occurs is a constant. Wien's law has wider significance than this, however Dividing Planck's law by T^5 results in

$$R_\lambda / T^5 = 37{,}410/(\lambda T)^5 (e^{14{,}380/\lambda T} - 1)$$

On the right-hand side of this equation the wavelength and temperature always appear multiplying each other. This means that only one curve is needed to express Planck's law, for all temperatures, if a graph is used in which the radiation energy divided by the fifth power of the temperature is plotted versus the product of wavelength and temperature. This is illustrated in Fig 4. It is helpful here to measure temperature in thousands of degrees. For comparison, the Rayleigh-Jeans law is also illustrated.

Wien's displacement law is helpful in determining the temperature of hot bodies. If λ_{max} can be measured, the temperature is immediately obtained from the displacement law. This is how an astronomer measures the temperature of a star. As an example, the radiation from the sun's surface has a maximum in the green region of the energy spectrum, in the vicinity of $\lambda = 0.5\ \mu$ (see Fig. 3). From Wien's law the surface temperature of the sun must then be about 6000°K.

Stefan-Boltzmann law. This law states that the total energy radiated from a hot body increases with the fourth power of the temperature of the body. This law can be derived from Planck's law by the process of integration and is expressed mathematically thus

$$R_T = 5.669 \times 10^{-12} T^4 \quad \text{(watts/cm}^2\text{)}$$

where R_T is the total amount of energy radiated from a black body. When R_T is multiplied by the total emissivity, the total energy radiated from a real heat radiator is obtained.

The rapid increase in heat radiation with temperature is quite evident from the Stefan-Boltzmann law If the absolute temperature is doubled, say from 273°K to 546°K (32° to 524°F), then the energy radiated increases sixteenfold. Thus, the attainment of very high temperatures requires large amounts of energy to overcome the loss of energy by heat radiation. Temperatures greater than 3×10^7 °K are encountered in a hydrogen bomb explosion. Such temperatures are 100,000 times higher than room temperature. Therefore, the energy radiated by 1 cm² of a substance at this high temperature will be 1×10^{20} times as much as that radiated at room temperature by the same substance. This energy, if radiated by a black body, would boil 2×10^7 tons of ice water in 1 sec.

Temperature determination. The apparent temperature of a real heat radiator can be determined by comparison with a black body whose temperature is known. This is done in any one of three customary ways, based upon the radiation laws described.

1 The radiation temperature of a surface is the temperature of the black body which radiates the same total energy per unit area in 1 sec as does the surface. This temperature is based upon the Stefan-Boltzmann law.

2. The brightness temperature of a surface is the

and the brightness temperature thus measured can be used in conjunction with the spectral emissivity $\epsilon_{0.655}$ and Wien's radiation law to calculate the true temperature of the body.

3. The color temperature of a surface is the temperature of a black body whose radiation has the same or approximately the same energy distribution as the surface. This is illustrated in Fig. 1a where it is seen that the xenon high-pressure discharge has a color temperature approximating 6000°K. For a gray body, that is, a body whose emissivity does not vary throughout the spectrum, the color temperature and true temperature are

Heat and Temperature Measurement, 1950; A. G. Worthing and D. Halliday, *Heat*, 1948.

Heat transfer

Heat, a form of kinetic energy, is transferred in three ways: conduction, convection, and radiation. Heat can be transferred only if a temperature difference exists, and then only in the direction of decreasing temperature. Beyond this, the mechanisms and laws governing each of these ways are quite different. This article gives introductory information on the three types of heat transfer (also called thermal transfer) and on important industrial devices called heat exchangers. For extended discussions of these topics, *see* CONDUCTION (HEAT); CONVECTION (HEAT); HEAT EXCHANGER; HEAT RADIATION. *See also* HEAT.

Conduction. Heat conduction involves the transfer of heat from one molecule to an adjacent one as an inelastic impact in the case of fluids, as oscillations in solid nonconductors of electricity, and as motions of electrons in conducting solids such as metals. Heat flows by conduction from the soldering iron to the work, through the brick wall of a furnace, through the wall of a house, or through the wall of a cooking utensil. Conduction is the only mechanism for the transfer of heat through an opaque solid. Some heat may be transferred through transparent solids such as glass, quartz, and certain plastics, by radiation. In fluids, the conduction is supplemented by convection, and if the fluid is transparent, by radiation.

The conductivities of materials vary widely, being greatest for metals, less for nonmetals, still less for liquids, and least for gases. Any material which has a low conductivity may be considered to be an insulator. Solids which have a large conductivity may be used as insulators if they are distributed in the form of granules or powder, as fibers, or as a foam. This increases the length of path for heat flow and at the same time reduces the effective cross-sectional area, both of which decrease the heat flow. Mineral wool, glass fiber, diatomaceous earth, glass foam, Styrofoam, corkboard, Celotex, and magnesia are all examples of such materials.

Convection. Heat convection involves the transfer of heat by the mixing of molecules of a fluid with the body of the fluid after they have either gained or lost heat by intimate contact with a hot or cold surface. The transfer of heat at the hot or cold surface is by conduction. For this reason, heat transfer by convection *cannot* occur without conduction. The motion of the fluid to bring about mixing may be entirely due to differences in density resulting from temperature differences, as in natural convection, or it may be brought about by [illegible] fact, the heat from the fire in the furnace heating the hot water or steam is transferred to the boiler wall by convection, and the hot water or steam transfers heat from the boiler to the radiator by convection. Iced tea is cooled and soup heated by convection.

Radiation. Solid material, regardless of temperature, emits radiations in all directions. These radiations may be, to varying degrees, absorbed, reflected, or transmitted. The net energy transferred by radiation is equal to the difference between the radiations emitted and those absorbed.

The radiations from solids form a continuous spectrum of considerable width, increasing in intensity from a minimum at a short wavelength through a maximum and then decreasing to a minimum at a long wavelength. As the temperature of the object is increased, the entire emitted spectrum decreases in wavelength. As the temperature of an iron bar, for example, is raised to about 1000°F, the radiations become visible as a dark red glow. As the temperature is increased further, the intensity of the radiation increases and the color becomes more blue. This process is quite apparent in the filament of a light bulb. When the bulb is operated at less than normal voltage, the light appears quite red. As the voltage is increased, the filament temperature increases and the light progressively appears more blue.

Liquids and gases only partially absorb or emit these radiations, and do so in a selective fashion. Many liquids, especially organic liquids, have selective absorption bands in the infrared and ultraviolet regions. *See* ABSORPTION (ELECTROMAGNETIC RADIATION).

Transfer of energy by radiation is unique in that no conducting substance is necessary, as with conduction and convection. It is this unique property that makes possible the transfer of large amounts of energy from the sun to the earth, or the transfer of heat from a radiant heater in the home. It is the ready transfer of heat by radiation from a California orange grove to outer space on a clear night that sometimes results in a frost. The presence of a shield of clouds will tend to prevent this loss of heat and often prevent the frost. By means of heat lamps and gold-plated reflectors, heat may be transferred deep into the layer of enamel on a car body, with resultant hardening of the enamel from the inside out. It is also the transfer over great distance of quantities of radiant energy that makes the atomic bomb so destructive.

Design considerations. By utilizing a knowledge of the principles governing the three methods of heat transfer and by a proper selection and fabrication of materials, the designer attempts to obtain the heat flow required for his purposes. This may involve the flow of large amounts of heat to some point in a process or the reduction in flow in others. It is possible to employ all three methods of heat transfer in one process. In fact, all three methods operate in processes that are commonplace. In summer, the roof on a house becomes quite hot because of radiation from the sun, even though the wind is carrying some of the heat away

by convection. Conduction carries the heat through the roof where it is distributed to the attic by convection. The prudent householder attempts to reduce the heat that enters the rooms beneath by reducing the heat that is absorbed in the roof by painting the roof white. He may apply insulation to the underside of the roof to reduce the flow of heat through the roof. Further, heated air in the attic may be vented through louvers in the roof.

Heat transferred by convection may be transferred as heat of the convecting fluid or, if a phase change is involved, as latent heat of vaporization, solidification, sublimation, or crystallization. The human body can be cooled to less than ambient temperature by evaporation of sweat from the skin. Dry ice absorbs heat by sublimating the carbon dioxide. Heat extracted from the products of combustion in the boiler flows through the gas film and the metal tube wall and converts the water inside the tube to steam, all without greatly changing the temperature of the water.

Heat exchangers. In industry it is generally desired to extract heat from one fluid stream and add it to another. Devices used for this purpose have passages for each of the two streams separated by a heat-exchange surface in the form of plates or tubes and are known as heat exchangers. Needless to say, the automobile radiator, the hot-water heater, the steam or hot-water radiator in a house, the steam boiler, the condenser and evaporator on

fineries, and chemical plants, two commonly used heat exchangers are the tube-and-shell and the double-pipe exchangers. The first consists of a bundle of tubes inside a cylindrical shell. One fluid flows inside the tubes and the other between the tubes and the shell. The double-pipe type consists of one tube inside another, one fluid flowing inside the inner tube and the other flowing in the annular space between tubes. In both cases, the tube walls serve as the heat-exchange surface. Heat exchangers consisting of spaced flat plates with the hot and cold fluids flowing between alternate plates are also in use. Each of these exchangers essentially depends upon convection heat flow through a film on each side of the heat-exchange surface and conduction through the surface. Countless special modifications, often also utilizing radiation for heat transfer, are in use in industry.

In these exchangers, the fluid streams may flow parallel concurrently, or in mixed flow. In most cases, the temperatures of the various streams remain essentially constant at a given point, and the process is said to be a steady-state process. As the streams move through the exchangers, unless there is a phase change, the fluids are continuously changing in temperature, and the temperature gradient from one stream to the other may be continuously varying. To determine the amount of surface needed for a given process, the designer must evaluate the effective temperature gradient for the particular condition and exchanger.

With extremely high temperatures, or with gas streams carrying suspended solids, the use of conventional heat exchangers becomes impractical. Under these conditions, the transfer of heat from one stream to another becomes more economical by the alternate heating and cooling of refractory solids or by checkerwork as in the blast-furnace hot stove, in the glass-furnace regenerator, or in the Royster stove. At lower temperatures, metal packing is frequently employed, as in the Ljungstrom preheaters or in regenerators for liquid-air production. In petroleum refining and in the metallurgical industry, exchangers are being employed in which one or more of the streams are fluidized beds of solids, the large area of the solids tending to produce very high rates of heat exchange. In some of these devices and also in nuclear power reactors, large quantities of heat are being generated in the exchangers. Here one of the principal problems involves the rapid removal of this heat before the temperature rises to the point where the equipment is damaged or destroyed.

Often the heating or cooling of a body is desired. In this case, the body representing the second stream does not remain at constant temperature, the heat being transferred representing a change in the heat content of the body. Such a process is

brick in a kiln; and the calcination of gypsum are examples of this type of process. [R.H.L.]

Bibliography: G. M. Dusinberre, *Numerical Analysis of Heat Flow*, 1949; H. C. Hottel and E. S. Cohen, Radiant heat exchange in gas filled enclosure, *A.I.Ch.E. Journal*, 4(1):3–14, 1958; M. Jakob, *Heat Transfer*, vol. 1, 1949; D. Q. Kern, *Process Heat Transfer*, 1950; J. G. Knutson and D. L. Katz, *Fluid Dynamics and Heat Transfer*, 1958; B. Lubarsky and S. J. Kaufman, Review of experimental investigations of liquid-metal heat transfer, *Natl. Advisory Comm. Aeronaut. Tech. Notes*, 3336:115, 1955; W. H. McAdams, *Heat Transmission*, 3d ed., 1954.

Heating, comfort

The maintenance of the temperature in a closed volume, such as a home, office, or factory, at a comfortable level during periods of low outside temperature. Two principal factors determine the amount of heat required to maintain a comfortable *inside* temperature: difference between inside and outside temperatures, and the ease with which heat can flow out through the enclosure. *See* CONDUCTION (HEAT).

Heating load. The first step in planning a heat-

outside and inside space temperatures and upon the heat transfer coefficients of the surrounding structural members.

Outside and inside design temperatures are first selected. Ideally, a heating system should maintain the desired inside temperature under the most severe weather conditions. Economically, however, the lowest outside temperature on record for a locality is seldom used. The design temperature selected depends upon the heat capacity of the structure, amount of insulation, wind exposure, proportion of heat loss due to infiltration or ventilation, nature and time of occupancy or use of the space, the difference between daily maximum and minimum temperatures, and other factors. Usually the outside design temperature used is the lowest temperature that has occurred once in 13 years.

The selected inside design temperature depends upon the use and occupancy of the space. Generally it is between 66 and 75°F.

The total heat loss from a space consists of losses through windows and doors, walls or partitions, ceiling or roof, and floor, plus air leakage or ventilation. All items but the last are calculated from

$$H_l = UA(t_i - t_o)$$

Where heat loss H_l is in British thermal units per hour, U is over-all coefficient of heat transmission from inside to outside air in Btu/(hr) (ft²) (°F), A is inside surface area in ft², t_i is inside design temperature, and t_o is outside design temperature

materials and types of construction from heating guides and handbooks.

The heating engineer should work with the architect and building engineer on the economics of the completed structure. Consideration should be given to the use of double glass or storm sash in areas where outside design temperature is 10°F or lower. Heat loss through windows and doors can be more than halved and comfort considerably improved with double glazing. Insulation in exposed walls, ceilings, and around the edges of the ground slab can usually *reduce* local heat loss 50–75%. Table 1 compares two typical dwellings. The 43% reduction in heat loss of the insulated house produces a worthwhile decrease in the cost of the heating plant and its operation. Building the house tight *reduces* the normally large heat loss due to infiltration of outside air.

Insulation and vapor barrier. Good insulating material has air cells or several reflective surfaces (*see* INSULATION, HEAT). A sound vapor barrier should be used with or in addition to insulation, or serious trouble may result. Outdoor air or any air at subfreezing temperatures is comparatively dry, and the colder it is the drier it can be. Air inside a space where moisture has been added from cooking, washing, drying, or humidifying will have a

Table 1. Effectiveness of double glass and insulation*

Heat-loss members	Area, ft²	Heat loss, Btu/hr: With single-glass weather-stripped windows and doors	Heat loss, Btu/hr: With double-glass windows, storm doors, and 2 in. wall insulation
Windows and doors	439	39,600	15 800
Walls	1,952	32 800	14 100
Ceiling	900	5 800	5 800
Infiltration		20 800	20 800
Total heat loss		99,000	56 500
Duct loss in basement & walls (20% of total loss)		19 800	11 300
Total required furnace output		118,800	67 800

* Data are for two-story house with basement in St. Louis

much higher vapor pressure than cold outdoor air Therefore, moisture in vapor form will pass from the high vapor pressure space to the lower pressure space and will readily pass through most building materials. When this moisture reaches a subfreezing temperature in the structure, it may condense and freeze at this point. When the structure is later warmed, this moisture will thaw and soak the building material, which may be harmful For example, in a house that might have 4 in. or more of mineral wool insulation in the attic floor, moisture can penetrate up through the second floor ceiling and freeze in the attic when the temperature there is below freezing. When a warm day comes, the ice will melt and can ruin the second floor ceiling Ventilating the attic helps because the dry outdoor air readily absorbs the moisture before it condenses on the surfaces. Installing a vapor barrier in insulated outside walls is recommended, preferably on the room side of the insulation. Good vapor barriers include asphalt-impregnated paper, metal foil, and some plastic-coated papers. The joints should be sealed to be most effective.

Infiltration. In Table 1, the loss due to infiltration is large. It is the most difficult item to estimate accurately and depends upon how well the house is built. If a masonry or brick veneer house is *not* well calked or if the windows are not tightly fitted and weather stripped, this loss can be quite large. Sometimes, infiltration is figured more accurately by measuring the length of crack around windows and doors.

Illustrative quantities of air leakage for various types of window construction are shown in Table 2. The figures given are in cubic feet of air per foot of crack per hour.

Design. Before a heating system can be designed, it is necessary to estimate the heating load for each room so that the proper amount of radiation or the proper size of supply air outlets can be selected and the connecting pipe or duct work designed. *See* CENTRAL HEATING; HOT-WATER HEAT-

Table 2. Infiltration loss with 15-mph outside wind

Building item	Infiltration, $ft^3/(ft)(hr)$
Double-hung unlocked wood sash windows of average tightness, nonweather-stripped including wood frame leakage	39
Same window, weather-stripped	24
Same window poorly fitted, nonweather-stripped	111
Same window poorly fitted, weather-stripped	34
Double-hung metal windows unlocked, non-weather-stripped	74
Same window, weather-stripped	32
Residential metal casement, $\frac{1}{64}$-in. crack	33
Residential metal casement, $\frac{1}{32}$-in. crack	52

ING SYSTEM; RADIATOR; REGISTER, AIR; STEAM HEATING; WARM AIR HEATING SYSTEM.

Heat is released into the space by electric lights and equipment, by machines, and by people. Credit for these in reducing the size of the heating system can be given only to the extent that the equipment is in use continually or if forced ventilation, which may be a big heat load factor, is not used when these items are not giving off heat, as in a factory. When these internal heat gain items are large, it may be advisable to estimate the heat requirements at different times during a design day under different load conditions to be certain that the inside temperatures can be maintained at the desired level.

Cost of operation. Design and selection of a heating system should include operating costs. The quantity of fuel required for an average heating season may be calculated from

$$F = \frac{Q \times 24 \times \mathrm{DD}}{(t_i - t_o) \times \mathrm{Eff} \times H}$$

where F = annual fuel quantity, same units as H
Q = total heat loss, Btu/hr
t_i = inside design temperature, °F
t_o = outside design temperature, °F
Eff = efficiency of total heating system (not just the furnace) as a decimal
H = heating value of fuel
DD = degree-days for the locality for 65°F base which is the sum of 65 minus each day's mean temperature for all the days of the year (*see* DEGREE-DAY).

If a gas furnace is used for the insulated house of Table 1, the annual fuel consumption would be

$$F = \frac{56{,}500 \times 24 \times 4699}{75 - (-5) \times 0.80 \times 1050} = 94{,}800 \text{ ft}^3$$

For a 5F°, 6- to 8-hour night setback, this consumption would be reduced by about 5%. *See* COMFORT CONTROL. [G.B P.]

Bibliography: Am. Soc. Heating, Refrigerating, Air Conditioning Engrs., *Heating Ventilating Air Conditioning Guide*, 1959; C. Strock (ed.), *Handbook of Air Conditioning Heating and Ventilating*, 1959.

Heating, electric

Methods of converting electric energy to heat energy by resisting the free flow of electric current. Electric heating has no upper limit to the obtainable temperature except for the materials used. It has the advantage of electrical temperature control and provides uniform heating.

When comparing economics of electric heating with other methods, the total costs must be considered, not just a comparison of the costs of equal Btu hours from electricity and fuels. The cost per unit of manufactured product is the true measure. This involves labor, quality, time, cleanliness, safety, and maintenance costs. Electric heat can usually excel in all of these.

Types of electric heaters. There are four major methods of electric heating.

1. Resistance heaters produce heat by passing an electric current through a resistance. Resistance heaters have an inherent efficiency of 100% in converting electric energy into heat. A high proportion of this heat can be transferred to the work material by conduction and by radiation. *See* RESISTANCE HEATING.

2. Dielectric heaters use currents of high frequency which generate heat by dielectric hysteresis within the body of a nominally nonconducting material. The power factor varies but, due to inherent losses, can never approach unity. *See* DIELECTRIC HEATING.

3. Induction heaters produce heat by means of a periodically varying electromagnetic field within the body of a nominally conducting material. The power factor can never approach unity, because of inherent reactances. *See* INDUCTION HEATING.

4. Electric-arc heating is really a form of resistance heating in which a bridge of vapor and gas carries an electric current between electrodes. The arc has a property of resistance. Both electrodes may be of carbon, or one may be the conducting work material in a furnace. *See* ARC HEATING.

Thermal problems. Electric heating differs radically from other methods that have a constant temperature. The heat-energy input of those methods depends on the temperature difference, which decreases as the work temperature rises. The heat input decreases as the work temperature increases.

Electric heating has a constant heat-energy input (for a given voltage), and the surface temperature of the heater rises to compensate for work temperature rise. Therefore, some control function is necessary to prevent overheating. To avoid damage, a suitable heat density per unit of heating surface must be maintained. With fluids, heat is distributed by natural thermal or forced convection currents. This helps heat transfer by wiping away surface films, and greater heat density can be tolerated. Heat capacity and mass play large parts in satisfactory heating. In large masses, such as tanks of oil or water, the temperature changes slowly and seldom presents a serious control prob-

lem. With small masses, uneven heat zones, or variable heating time, the control of the input heat becomes more critical. Heat control is required when heating such materials as gas vapor and paper Moisture content also affects the heat requirements.

The heater rating can be found from the heat required for the process. This must include the heat required to heat the material and its container, any heat of fusion or vaporization in the process, and the heat losses. The heat required for the material and its container can be found from the weight, the specific heat, and the temperature rise of the material and container. The total heat required for the process is the sum of all the pertinent factors and can be put in units of kilowatt-hours. The number of kilowatt-hours divided by the hours allowed for the process will then give the required rating of the heater in kilowatts. The heater rating therefore depends on the time allowed for the process. A longer time will permit use of a heater of lower rating and result in reduced costs of equipment and operation.

General design features. All electrical parts must be well protected from contact by operators, work materials, and moisture. Terminals must be enclosed within suitable boxes, away from the high heat zone, to protect the power supply cables. Repairs and replacements should be possible without tearing off heat insulations.

Resistance heaters are often enclosed in pipes or tubes suitable for immersion or for exposure to difficult external conditions. Indirect heating is done by circulating a heat transfer medium, such as special oil or Dowtherm (liquid or vapor), through jacketed vessels. This permits closer control of heating surface temperature than is possible with direct heating.

Some conducting materials can be heated by passing electric current through them, as is done in the reduction of aluminum. Some conducting liquids can be heated by passing an electric current between immersed electrodes. Heat is produced by the electrical resistance of the liquid path.

The supply of necessary electric power for large heating installations necessitates consultation with the utility company. The demand, the power factor of the load, and the load factor all affect the power rates. Large direct-current or single-phase alternating-current loads should be avoided. Polyphase power at 440–550 volts permits lower current and reduced costs. *See* FURNACE CONSTRUCTION.

[L.P.HY.]

Bibliography: Edison Electric Institute, *Power Sales Manual: Induction and Dielectric Heating*, 1949; L. P. Hynes, Industrial electric resistance heating, *AIEE Trans.* 67:1359–1361, 1948; A. J. Johnson and G. H. Auth (eds.), *Fuels and Combustion Handbook*, 1951; H. Pender and W. A. Del Mar, *Electrical Engineers' Handbook*, 4th ed., vol. 1, 1949.

Heat-treatment (metals and alloys)

A process which may include several steps involving the heating and cooling of a metal or alloy. The purpose of heat-treating is to obtain certain desired properties in the product, such as ductility, hardness, or toughness. The metal is not melted during treating, and both the rate and conditions of heating and cooling, as well as the time that the material is held at any temperature, must be carefully regulated.

HEAT–TREATMENT OF NONFERROUS METALS AND ALLOYS

Because common nonferrous metals do not show allotropic transformations, it is generally not possible to harden them by a simple heating and quenching treatment as with steels. Also, in contrast to steels, it is generally not possible to produce grain refinement in nonferrous metals by heat-treatment alone.

Annealing. This process involves heating the metal or alloy to a temperature at which grains deformed by cold-work recrystallize to produce new grains. The annealing, or recrystallization, temperature depends on the material and on the degree of cold-working. As the deformation caused by cold-working increases, the recrystallization or annealing temperature decreases. During recrystallization the hardening effect of cold-work is eliminated and the ductility of the material increases. As the annealing temperature is increased the size of the newly formed grains increases; the annealing temperature that produces the most desirable mechanical properties is the lowest temperature that results in complete recrystallization. An increase in the time that the metal or alloy is held at the annealing temperature results in an increase in the grain size of the material. However, the influence of annealing time is much less important than that of annealing temperature. Once the material has recrystallized at the annealing temperature, its properties are unaffected by the rate of cooling, provided that no phase change occurs between the annealing temperature and atmospheric temperature. Most nonferrous materials do not undergo a phase change on cooling. The range of annealing temperatures for copper-base alloys is 550–800°F, 700–1500°F, for aluminum-base alloys 550–800°F, and for nickel-base alloys 1500–2200°F. Lead and its alloys recrystallize at atmospheric temperatures and therefore cannot be hardened by cold working; such materials are said to be self-annealing. In order to prevent the formation of an oxide scale on heating, some metals and alloys are annealed by heating in a protective or nonoxidizing atmosphere. Such an atmosphere may be made by the partial combustion of hydrocarbons, or it may consist of cracked ammonia or a nonoxidizing gas such as argon. Some of the highly reactive metals such as titanium and zirconium may be annealed in a vacuum.

Stress relieving. This operation consists of heating the metal to a temperature below the recrystallization or annealing temperature but high enough to relieve residual stresses introduced by plastic deformation. It is a treatment usually applied to finished parts.

Heating for homogenization. This is performed to decrease, or eliminate, segregation, that is, nonuniformity of composition. Segregation occurs during freezing of alloys. If the alloy is heated to a temperature not far below its freezing temperature and held at that temperature for a long time, interdiffusion of the alloy constituents will tend to eliminate segregation. Such homogenization treatment is frequently called soaking.

Hardening. Certain nonferrous alloys can be hardened by a precipitation-hardening treatment, sometimes spoken of as an age-hardening treatment. To be amenable to such treatment the solubility of an element or compound in the alloy must increase with increasing temperature, and the amount of this constituent must exceed its solid-solubility limit at low temperatures. Heat-treatment for hardening consists of heating the alloy to a temperature not much below the solidification temperature, holding it at this temperature for a time sufficient for dissolution of the phase that precipitates at lower temperatures, and then cooling it rapidly enough to prevent precipitation of the second phase on cooling. On subsequent heating to a lower temperature, or by simply letting certain alloys remain at atmospheric temperature, some of the second phase will precipitate from the supersaturated solid solution. At the proper reheating, or aging, temperature the precipitated particles will be extremely fine and will harden and strengthen the alloy. The first treatment is spoken of as the solution treatment or solution anneal. The second treatment is termed the aging treatment. A well-known precipitation-hardenable aluminum alloy contains 4% copper, 0.5% magnesium, and 0.5% manganese. It is solution-treated by heating it to 925–950°F, usually in a salt bath, holding it at this temperature for as long as 1–2 hours, and then quenching it in water. On standing at atmospheric temperature the hardness and strength will gradually increase, maximum values being reached in 5–6 days.

Copper containing 2% beryllium can be hardened by solution annealing at 1400°F and then aging at 600°F for 3 hours.

Some of the less common nonferrous metals, such as titanium [illegible]

temperature phase is stable. On heating through the transformation temperature the original grains disappear and new grains of the high-temperature phase appear. On cooling the reverse action takes place. Grain refinement may result from these transformations, but it is usually necessary to heat the material slightly above the transformation temperature. This must be done because, as the temperature is increased, the grains of the high-temperature phase grow and the size of the ultimate grains of the low-temperature phase seems to be approximately that of the grains of the high-temperature phase from which they were formed. Some nonferrous alloys in which the basis metal exhibits allotropic transformations can be hardened and tempered just as steels are hardened and tempered. The principles and methods involved are similar to those involved with iron-carbon alloys. *See* ALLOY.

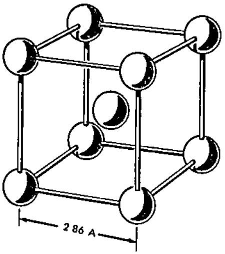
[J.L.CG.]

HEAT-TREATMENT OF STEELS

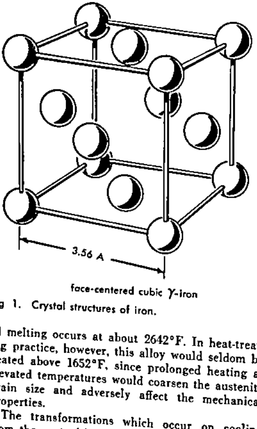
Most of the heat-treatments of steels involve the transformation of iron from one crystal structure to another. Below 1670°F iron exists in a body-centered cubic structure, α-iron, illustrated in Fig. 1. In the range 1670–2552°F, iron has a face-centered cubic structure, γ-iron, also shown in Fig. 1. Between 2550°F and the melting point, 2802°F, iron reverts to the body-centered form, δ-iron, which has the same structure as α-iron. Practical alloys are not single crystals with these structures but consist of many abutting crystals, or grains, more or less randomly oriented with respect to each other. Many of the heat-treatments described in the following sections are designed to change the grain size or the crystal structure.

General principles. The addition of carbon to iron influences the transformation from one form to another, and the resultant structures are summarized in the iron-carbon phase diagram shown in Fig. 2. This diagram indicates the microstructures, or phases, which are found for a given composition on prolonged heating at each temperature. Iron and carbon form an intermetallic compound, cementite, Fe_3C, and the phases found in iron-carbon alloys are α-iron with carbon in solution (ferrite), γ-iron with carbon in solution (austenite), and cementite. Cementite is not the most stable form for excess carbon in iron, and true equilibrium is achieved on the formation of graphite after prolonged heating. Graphite is observed in the common grades of cast iron, which contain 2–4% carbon and a high silicon content to aid graphitization. The carbon atoms in solution in the iron occupy interstitial positions between the iron atoms and expand the lattice slightly. The maximum solubility of carbon in α-iron is 0.025 wt % at 1333°F, and in γ-iron, 2.0 wt % at 2066°F.

The changes that occur on slowly heating a typical steel may be illustrated with the aid of the phase diagram (Fig. 2). A steel containing 0.40% carbon, for example, will consist of ferrite and cementite at room temperature. On heating above the A_1 critical temperature 1333°F, austenite forms, and ferrite and austenite coexist until the A_3 critical temperature is reached. Above A_3 the structure is entirely austenitic. On heating above the A_3 temperature, few changes are observed un-



body-centered cubic α-iron



face-centered cubic γ-iron

Fig 1. Crystal structures of iron.

til melting occurs at about 2642°F. In heat-treating practice, however, this alloy would seldom be heated above 1652°F, since prolonged heating at elevated temperatures would coarsen the austenite grain size and adversely affect the mechanical properties.

The transformations which occur on cooling from the austenitic region depend on the cooling rate. For a slow cooling rate, that is, furnace cooling, the sequence is the reverse of that described above. Below A_3 ferrite is first formed and this co-exists with [illegible] cementite and ferrite form alternate platelets in an aggregate called pearlite. The average composition of the pearlite colonies is 0.80% carbon and a steel of this composition (eutectoid) would consist entirely of pearlite. The microstructure of a 0.40% carbon steel, obtained by polishing and etching and examining the surface at 1000 magnification, is shown in Fig. 3. About half of the structure is pearlite and the remainder ferrite.

Annealing and normalizing. A typical annealing treatment for steels containing less than 0.80% carbon (that is, hypoeutectoid) involves heating to 45–90°F above the A_3 temperature and furnace cooling. Normalizing treatments are similar but involve air cooling from somewhat higher temperatures. Several objectives may be accomplished by these treatments. On heating through the A_1 temperature, austenite is first formed. The first austenite grains arise from nuclei formed at the junctions of two or more ferrite, or pearlite, boundaries and these first austenite grains have a rather small grain size. The remaining austenite is formed on

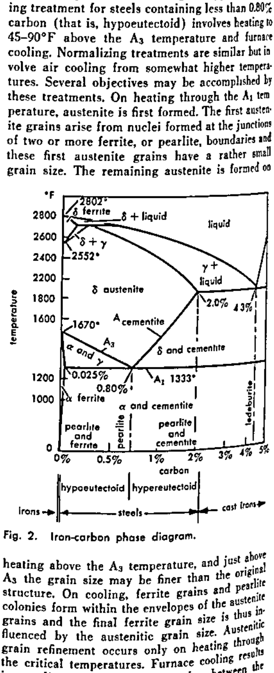


Fig. 2. Iron-carbon phase diagram.

heating above the A_3 temperature, and just above A_3 the grain size may be finer than the original structure. On cooling, ferrite grains and pearlite colonies form within the envelopes of the austenite grains and the final ferrite grain size is thus influenced by the austenitic grain size. Austenitic grain refinement occurs only on heating through the critical temperatures. Furnace cooling results in pearlite with a coarser spacing between the lamellar plates than air cooling. The finer interlamellar spacing in the pearlite of normalized steels produces a somewhat harder and stronger material, but the slower cooling of the annealing treatment results in lower residual stresses. The principal objectives in annealing and normalizing treatments are grain refinement, stress relief, and some chemical homogenization.

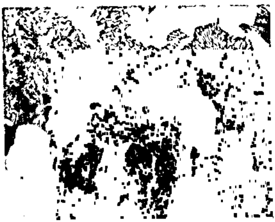


Fig. 3. Microstructure of annealed 0.40% carbon steel, × 1000. White regions are ferrite; lamellar regions are pearlite.

Hypereutectoid steels (containing more than 0.80% carbon) are treated somewhat differently. Above the A_1 temperature the structure consists of austenite and cementite, and above the A_{cm} temperature, the structure is entirely austenitic. On annealing these steels, they are cooled slowly through the A_{cm} temperature, and in some instances heated and cooled several times through A_1 in order to promote the formation of spheroidized carbides rather than pearlite. Spheroidization results in an improvement in the machinability of these steels.

Annealing and normalizing are commonly used as stress-relieving treatments. In complex weldments and castings, however, stress relief may be accomplished by heating to a temperature below A_1, 1100–1200°F, and slow cooling or air cooling. Residual stresses may be reduced considerably by these treatments with somewhat less danger of distorting the shape of the piece.

Isothermal transformation diagrams. The influence of cooling rate on the transformation of austenite may be illustrated by considering the isothermal transformation diagram for a particular steel, for instance, a eutectoid steel in Fig. 4. This di...

temperature, and holding it at each temperature for ...

tur...

wit...

ten...

a complex aggregate of ferrite and carbide called bainite is formed. This structure is more commonly observed in some of the higher alloy steels. At 425°F, an entirely different transformation product, martensite, is formed from the austenite. This microstructure is basic to the hardening of steel and is discussed further below. However, it should be noted that the temperature at which martensite begins to form, the M_s temperature, is represented as a horizontal line on the diagram. This indicates that martensite formation cannot be avoided by rapid cooling. In contrast, pearlite requires a definite time to form and the pearlite structure can be suppressed by cooling rapidly enough to avoid the pearlite nose on the diagram. Various cooling rates are indicated on the isothermal transformation diagram showing how furnace cooling results in coarse pearlite, air cooling in fine pearlite, and rapid quenching in martensite.

Hardening and tempering. The hardening of steel is accomplished by heating into the austenitic region and cooling rapidly enough to avoid the formation of ferrite, pearlite, and bainite. Hypoeutectoid steels are quenched from above the A_3 temperature whereas hypereutectoid steels are quenched from a temperature just above A_1. Water and oil quenching are required for plain carbon steels. Alloy additions inhibit the formation of pearlite and other decomposition products and allow the use of slower cooling rates and the hardening of thick pieces. A steel containing 5% chromium and 1.5% carbon, for example, can be hardened by cooling in air.

Martensite is formed from austenite in the hardening operation. The horizontal lines below M_s in Fig. 4 show how the reaction proceeds on cooling to room temperature, and it should be noted that the transformation is independent of the cooling rate through the martensitic region. The reaction is not complete at room temperature and 5–10% of retained austenite may be present. Additional martensite formation occurs on continuing the cooling below room temperature. A variation of the cooling sequence, known as marquenching or martempering, involves quenching to a temperature in the vicinity of M_s, holding long enough to bring the piece to temperature equilibrium, and then air cooling to room temperature. This treatment avoids the introduction of severe quenching stresses and usually results in lower distortion and smaller residual stresses.

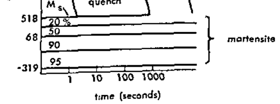
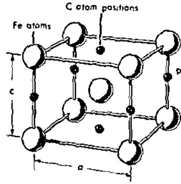


Fig. 4. Schematic isothermal transformation diagram for a eutectoid steel.



weight percent C	c(A)	a(A)
0	2 86	2 86
0 20	2 88	2 858
0 40	2 91	2 856
0 80	2 95	2 85

Fig. 5. Crystal structure of martensite.

Martensite is a body-centered tetragonal structure, shown in Fig. 5, with the carbon atoms in ordered positions in the lattice. The tetragonality and the hardness increase with the carbon content, the hardness reaching a plateau value of about 66 on the Rockwell C scale at 0.80% carbon.

Untreated martensite containing more than 0 10% carbon is brittle and a tempering operation is usually used to increase the ductility. On heating (or tempering) martensite, the structure decomposes in a series of complex reactions. For plain carbon steels and for heating times of about 1 hour these reactions may be considered in three stages: (1) at temperatures up to about 395°F, martensite decomposes to form ε-carbide, having a hexagonal structure with the approximate formula-tion Fe_2C, and a low-carbon martensite containing about 0.25% carbon; (2) in the range 395–600°F the retained austenite decomposes to bainite, and (3) in the range 395–1000°F, the ε-carbide and the low-carbon martensite decompose to form α-ferrite and cementite (Fe_3C). See IRON ALLOYS; METAL, MECHANICAL PROPERTIES OF; STEEL.

[B.L.A.]

Bibliography: American Society for Metals, *Metals Handbook*, 1948; supplement, 1955.

Heavy water

A compound of hydrogen and oxygen that contains a higher proportion of the hydrogen isotope of mass 2 (deuterium, D) than does naturally occurring water. Pure heavy water has the formula D_2O and is called deuterium oxide.

Uses. Heavy water is used in nuclear reactors as a moderator and as a coolant. As a moderator, it is about four times as efficient as graphite for slowing down fast neutrons. Heavy water, D_2O, also has a desirably low absorption cross section for neutron capture of 0.001 barns, as compared to 0.6 for H_2O. Heavy water reactors, planned or in operation, include several for power and research purposes. Heavy water has also been used extensively in chemical [illegible] tracer, me[illegible]

Proper[illegible] different [illegible] [illegible]incipal species being H_2O^{16}, HDO^{16}, H_2O^{17}, and H_2O^{18}. The deuterium content of natural water is about 0.015%, although it varies with location and history of the water. The O^{18} content is about 0.2%, and O^{17} content about 0.04%. Heavy water is water enriched in either deuterium or the heavier oxygen isotopes, and it differs from natural water in both physical and chemical properties. In most usage, however, heavy water refers to D_2O, that is, to the compound having two deuteriums in place of the hydrogens, and it is this species of heavy water which has been prepared in sizable quantities. Heavy water, D_2O, has a higher density, lower vapor pressure, and a higher melting point than ordinary water. Differences in select physical properties are shown in Table 1. The two waters differ only slightly in surface tension, but heavy water is the more viscous; for example, the ratio of viscosities $\eta D_2O/\eta H_2O = 1.23$ at 25°C. The refractive index of D_2O is appreciably smaller than that of H_2O, and this difference has been used as a basis for isotopic analysis of water.

Preparation. In the preparation of D_2O from natural water, advantage is taken of the differences in physical and chemical properties of the different isotopic species. Separation processes suitable for large-scale production of heavy water include (1) electrolysis, (2) distillation, (3) chemical exchange, and (4) dual-temperature exchange.

Prior to 1943, heavy water was produced commercially only by electrolysis. In this process, isotope fractionation occurs as the water is decomposed electrolytically. The deuterium content of the hydrogen produced is less than that of the water remaining in the electrolytic cell. Deuterium separation factors are high (3–8) and the process is readily adaptable to small-scale production. Although excellent for producing heavy water from partially concentrated material, electrolysis is not competitive economically with other processes when applied to natural water.

A method more suitable for natural water is one involving distillation. Since HDO and D_2O have lower vapor pressures than H_2O, they can be separated from H_2O by distillation. Separation factors vary with temperature and pressure, being 1.1 at 10°C (9.2 mm), 1.05+ at 50°C (92 5 mm), and 1.026 at 100°C (760 mm). Water distillation was used in three Manhattan District heavy water plants, in which deuterium was concentrated from 0.0143 atom % to about 90 atom %. Further concentration to 99.8 mole % D_2O was effected by electrolysis. The water distillation process has the advantage of producing D_2O directly from a raw

Table 1. Physical properties of H_2O and D_2O

	D_2O	H_2O
Boiling point, °C	101.4	100
Melting point, °C	3 8	0 0
Specific gravity (d_{25}^{25})	1.10775	1 0000
Temperature of maximum density, °C	11 23	3 98
Critical temperature, °C	371 5	374 2
Critical pressure, atm	218 6	218 5

Table 2. Separation factor for dual-temperature process

Exchange reaction	0°C	25°C	100°C
$H_2O(l) + HD$	4 69	3 87	2.69
$H_2O(g) + HD$	4.19	3 62	2 62
$H_2O(l) + HDS$	2 52	2 31	1.92
$H_2O(l)$ + isopropyl mercaptan (l)	2.17	2 02	1 75

material available in unlimited quantities. A more efficient method for concentrating deuterium is the liquid-hydrogen distillation process. Separation factors of about 1.4 at −250°C, and deuterium recovery of about 90% are reported.

The chemical exchange reactions that have been proposed for concentrating deuterium from natural water include the following:

$$H_2O + HD \rightleftharpoons HDO + H_2$$
$$H_2O + HDS \rightleftharpoons HDO + H_2S$$
$$H_2O + DCl \rightleftharpoons HDO + HCl$$

Other exchange processes studied are the isopropyl mercaptan–water system, the cyclohexane-benzene-hydrogen system, and various ammonia exchange processes. Of these, the water- (or steam-) hydrogen reaction is of special interest since it was the basis for the commercial production of heavy water at the Trail Plant in British Columbia. The separation factor for the steam-hydrogen exchange varies from about 4.2 at 0°C to 2.6 at 100°C. A catalyst such as platinum on activated carbon, or nickel-chromia, is needed. The exchange reaction is carried out in a series of towers. At the bottom of each tower, part of the concentrated water must be converted to hydrogen to provide gas for the exchange. In the Trail process, this is done by electrolysis, and the deuterium-free hydrogen is available for chemical use.

The need for electrolysis is avoided in the dual-temperature exchange process. This process effects deuterium concentration by utilizing the variation in separation factor with temperature for exchange reactions. The variation for four such reactions is shown in Table 2.

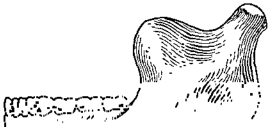
In the steam-hydrogen dual temperature process, hydrogen is recycled through two exchange towers, one at low temperature and the other at high. With proper choice of temperatures, deuterium is transferred from hydrogen to steam in the cold tower and from steam to hydrogen in the hot. The net result is a transfer of deuterium from steam in the hot tower to steam in the cold tower. The H_2O-HDS dual temperature process is similar except that H_2S is recycled between the hot and cold towers. *See* DEUTERIUM; ISOTOPE SEPARATION (STABLE ISOTOPES); LIQUID; REACTOR, NUCLEAR; TRITIUM; WATER. [I.K.]

Bibliography: L. M. Brown, A. S. Friedman, and C. W. Beckett, *Bibliography of Research on Deuterium and Tritium Compounds, NBS Circular* 562, 1956; I. Kirshenbaum, *Physical Properties and Analysis of Heavy Water*, 1951; G. M. Murphy (ed.), *Production of Heavy Water*, 1955. U.S. Patents 2.787.526 and 2.895,803.

Heidelberg man

An early fossil type known only from an isolated lower jaw. This was found at a depth of approximately 79 ft. in gravel laid down in an old bed of the Neckar River, at Mauer, 6 miles southeast of Heidelberg, Germany. Discovered by a workman in a commercial gravel pit on October 21, 1907, it was kept because of the watchfulness of O. Schoetensack of Heidelberg University. The jaw was not accompanied by stone implements. However, there

and marked by a very broad ascending ramus, which gives it a primitive appearance. The teeth, while robust and large-rooted, seem small for the jaw; they lack many primitive characteristics and bear a general but not complete resemblance to



The Heidelberg jaw. (*From M. F. Ashley Montagu, An Introduction to Physical Anthropology, 2d ed., Charles C Thomas, 1951*)

modern human teeth, and may be viewed as more advanced than the teeth of other early human fossils. The jaw has been named *Homo heidelbergensis*. *See* FOSSIL MAN. [W.W.H.]

Helicopter

An aircraft that sustains itself by motor-driven horizontal rotating blades (rotors) that accelerate

copter is the most successful rotary-wing vertical take-off and landing aircraft developed to date. Although many different arrangements of lifting rotors have been developed, the ones which have received practical application are dual rotors placed [illegible] with an anti-
[illegible] two counter-
[illegible]ted (Fig. 1).

In aeronautical development, the helicopter represents a logical progression from the autogiro by the use of the lifting rotor for forward propulsion and in turn serves as the link to the convertiplane which uses rotors or similar devices for vertical

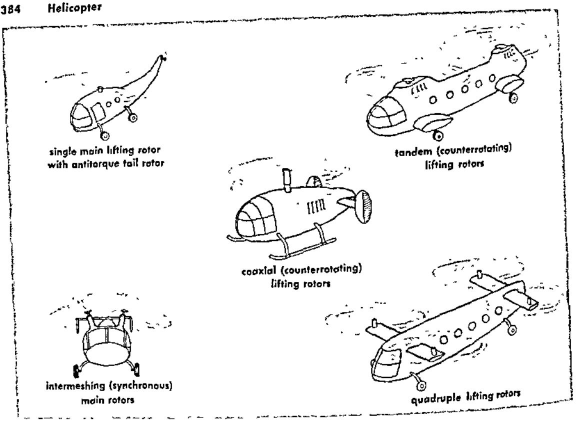


Fig. 1. Principal helicopter (rotor) configurations.

lift combined with other methods of forward propulsion in order to obtain increased speed.

Flight capabilities. The advantages of the helicopter over more traditional fixed-wing aircraft lie in its ability, by virtue of its vertical rising capability, to operate in very confined areas without the necessity of runways or other terminal facilities, coupled with its ability, by employing autorotation, to descend safely in the event of power failure. Although many slow-speed aircraft such as the autogiro have been designed, the helicopter remains uniquely capable of backing or sliding sideways into its confined landing area or hovering for prolonged periods over a fixed point.

These advantages are exploited in the uses of the helicopter in transportation as a vehicle for airport-to-city and interurban transportation as practiced by commercial operators in New York, San Francisco, Chicago, and other large metropolitan areas; its slow flight capabilities allow it to be used for forestry, pipeline, and ...

erating in the field with troops; its hovering capability makes it a spectacular rescue vehicle for uses as beach guards and as plane guards on aircraft

its slow forward flight speed as compared with that of other aeronautical devices and its vibration characteristics and mechanical features which tend to increase its operating cost over that of c

tional aeronautical equipment. These disadvantages have led to an intensive effort in recent years to seek a vehicle combining the slow flight characteristics and zero landing speed of the helicopter with the higher cruise speed of the aircraft. See AUTOGIRO; CONVERTIPLANE; VERTICAL TAKEOFF AND LANDING (VTOL).

Rotors. Vertical lift and forward propulsion of the helicopter are supplied by a lifting rotor (or rotors) operating on a vertical axis. The rotor con ... designed as airfoils and ro- ... t is varied by ... s are mounted

to the axis by means of pitch control (Fig 2). Because the rotors are operating at different air speeds as they progress around the disk, it is necessary to vary their lift effect in order to maintain uniform lift throughout the rotor disk as well as for control and stability. This is achieved by a hinge for the rotors and a variation of pitch in a cyclic direction, this same feature being employed to obtain the horizontal component of force required for forward motion. The pitch variation has the aerodynamic effect of tilting the rotor disk in the direction of flight, although physically the axis remains practically vertical. This variable pitch is known as cyclic or feathering action and requires a decrease in pitch on the blade traveling in the same direction as the helicopter (and hence at a higher relative air speed to fuselage) and an increase in pitch with the retreating blade traveling in opposition to the fuselage itself. One of the prime differences between the helicopter rotor systems em

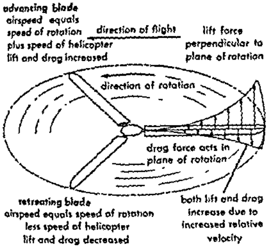


Fig. 2. Helicopter rotor aerodynamics.

ployed by various organizations and engineers lies in the method of achieving this variation in pitch in a cyclic sense without creating undesirable forces on the mast.

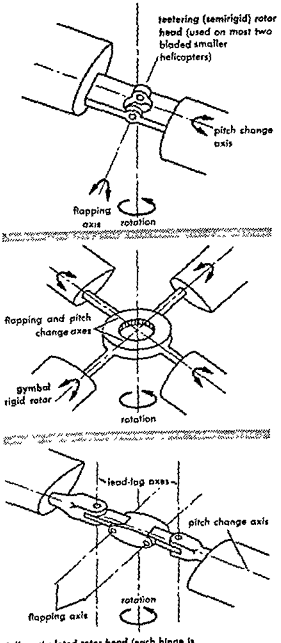
The first satisfactory solution to this problem by J. de la Cierva was the essential technical advance that made possible the successful flight of the autogiro (Fig. 3). The first solution to the problem lies in hinging the rotor blades in both the horizontal and vertical planes as well as varying the pitch of the blade-carrying socket as it progresses around the circle. This fully articulated system is typified by those used on the large helicopters manufactured by Sikorsky and Vertol. On smaller machines, it has been found possible to eliminate the hinges by allowing the rotor blade to bend, teetering the axis on two bladed rotors and, in at least one instance, four bladed rotors on a gimbal-mounting system. This approach is typified on the smaller helicopters of Bell and Hiller and in the four-bladed rotor version by the prototype models of Doman.

The basic theory of rotor aerodynamics follows propeller theory closely, although further refine-

difference arises from the fact that the helicopter disk is moving parallel (or nearly so) to the line of flight, and is therefore subject to a complex air inflow pattern compared to a propeller. Basically,

themselves are constructed of metal, wood, plastics, and other materials. The smaller machines have employed metal blades of various designs increasingly in recent times. In general, the rotor blade consists of a rather heavy spar, required not only for the normal bending and other stresses experienced in fixed wings, but also for the added centrifugal forces experienced in the rotating airfoil. The airfoil itself is constructed of wraparound designs using either small pockets or sheets of aluminum, plastic, or even lightweight steel and titanium to form the aerodynamic surfaces. Experimental efforts towards deicing and anti-icing of airfoil surfaces, including heating through ducts, have been made; however, most helicopter blades have no ice protection at the present time.

Power plants. Most in-service helicopters are powered by modified versions of standard aircraft reciprocating engines. Problems have been encountered because of the high power requirements imposed on these engines by helicopter operations



fully articulated rotor head (each hinge is restrained by dampers, limiting stops or other device)

Fig. 3. Rotor hub types.

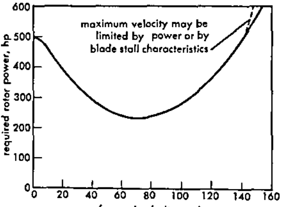


Fig. 4. Typical curve of required helicopter power showing high horsepower required at low and high flight speeds

(Fig. 4) Additional problems arise from the cooling and vibration, which lead to new design problems for the older aircraft engines available. Newer helicopter types coming into service are generally powered by gas turbine engines which, in view of their power characteristics and light weight per unit of power, offer tremendous advantages to the helicopter designer (*see* TURBINE PROPULSION). Most helicopters are driven by engines operating through reduction gears to the rotor.

The installation of a power plant at the helicopter rotor blade tip offers an effective means of powering the lifting rotor. The high centrifugal forces to which an engine [illegible]

only practicable solution to tip-powered helicopters to date has been in the pressure-jet principle in which a fuselage-mounted gas generator has pushed exhaust jets through the blade tips where it exhausts perpendicularly to drive the blade either cold or unaugmented, or hot by injecting fuel after the manner of a turbojet afterburner. Several smaller helicopters with tip-mounted ramjet, pulsejet, and other types of engine have been manufactured, but generally have not been accepted because of the high noise level of these types of engine. Experimental studies on the tip mounting of turbine engines have been conducted, and at least one experimenter has investigated the powering of the rotor blades by rotor-mounted propellers.

Flight characteristics. The helicopter achieves vertical flight by accelerating a mass of air downward through the rotor disk. At lower altitudes (up to one-half of the lifting rotor diameter) the effect of this downward velocity of air, termed downwash, results in an augmentation of the lift known by the term ground effect. In general, vertical lift above the region of ground effect requires the maximum horsepower of the helicopter, and most designs produced to date are seriously limited in this flight regime by the available installed horsepower. A slight amount of forward velocity reduces the power requirements for vertical lift considerably and rapidly, and consequently, most helicopters are operated with a forward climbing motion much steeper than but similar to that employed by fixed-wing aircraft.

In forward flight, the helicopter is limited in maximum speed by blade stall. Because the rotor blade has a constant still-air velocity, as the helicopter itself approaches this speed, the retreating blade tip approaches a net zero forward velocity. The ratio of forward speed of the helicopter to tip speed is termed advance ratio. Modern helicopters are operated at 50–60% advance ratio. Autogiros were customarily operated at 85–90% because of lesser vibratory effects from the high speeds on the compression blade side. Speed of rotation of the rotor is therefore a compromise between the high speed experienced on the advancing blade tips, which approach sonic velocity, and the retreating low speed, which limits the helicopter itself. In operating the helicopter one must avoid extended flight in the area of blade stall, in which excessive buffeting, loss of control, and serious flight problems can ensue.

All helicopters incorporate a one-way clutch system which allows the rotor blades to continue to rotate under the effects of inertia and autogiro-type flight in the event of engine failure. In effect, the flow of the air stream through the helicopter rotor disk is reversed. The force of gravity continues the operation of the rotor blade, thereby reducing the rate of descent and allowing a successful emergency landing. Because of the time required to effect this change and the necessity for other flight maneuvers, these characteristics led to the development of a restricted area of operations of the helicopter known as the deadman's curve, bounding that area of height and speed in which the helicopter should not be operated (Fig. 5).

Another interesting phenomenon of helicopter characteristics is the settling with power in which the helicopter, while rising vertically, creates a vortex effect of moving air in a vertical direction. This vortex tends to increase greatly the power required for hovering and true vertical flight. Earlier lower-powered helicopters frequently ran into serious problems in this flight regime, although

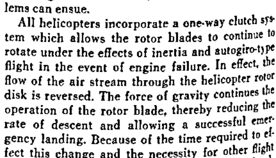


Fig. 5. Typical deadman's curve for operation of a helicopter. It should not be operated in shaded flight condition because recovery in the event of power failure would be difficult or impossible.

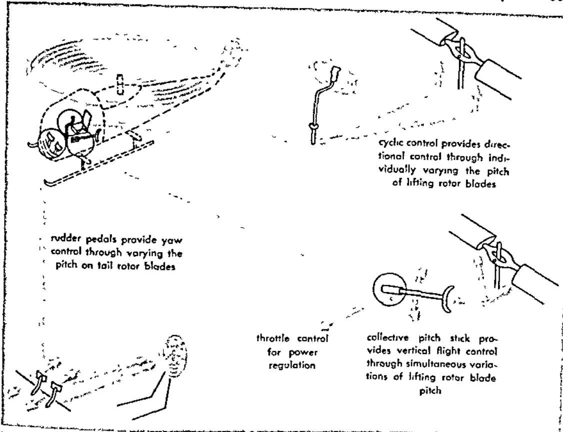


Fig. 6. Helicopter flight controls.

more modern and adequately powered helicopters do not normally experience this difficulty. Generally, however, helicopter pilots avoid the flight regimes which might result in settling with power or excessive power requirements in order to maintain proper altitude.

Operation. The flight controls of the helicopter are complex to allow full exploitation of its capability to move in all directions (Fig. 6). The primary control, termed the cyclic pitch, occupies a position between the pilot's knees similar to a control stick in a fixed-wing aircraft. The cyclic pitch control, in general, directs the aircraft in the same direction in which it is moved; if the stick is moved forward, the aircraft moves forward; if it is moved to the right the aircraft moves to the right, and so on.

The vertical direction of movement is controlled by the collective pitch, which moves the aircraft up if the stick is lifted and down if the stick is depressed. This control on most designs is placed by the pilot's left hand and while it may be locked temporarily with a friction lock, most of the time it must be held in the pilot's hand, particularly as a precaution against the necessity for sudden autorotation. The yaw or angular direction of the fuselage is controlled by rudder pedals similar to those employed by fixed-wing aircraft.

Because of the varying power requirements of a helicopter and its sensitivity to direction of movement, the throttle on the engine is constantly monitored on the helicopter. Normally, it is controlled by a motorcycle-type twist grip mounted on the left-hand collective pitch. A certain amount of coordination of this control is accomplished by automatically increasing the throttle setting as the collective pitch is lifted and decreasing it as it is lowered.

The basic flight characteristics of the helicopter, particularly the smaller, more agile ones, result in the interaction or feed-back effect among the various controls. Thus, for example, a forward move-

tude. This in turn has a lower power requirement and requires a slight reduction in throttle; because this decreases the torque effect of the main rotor, it will also require an adjustment of pedal pressure to maintain true direction.

to a higher degree of tension and fatigue than the pilot of a fixed-wing aircraft. Normal flight training requires approximately twice as much time as that of fixed-wing pilots.

Some success has been achieved in automation and electronic control of flight, and droned or pilotless helicopters have been demonstrated for

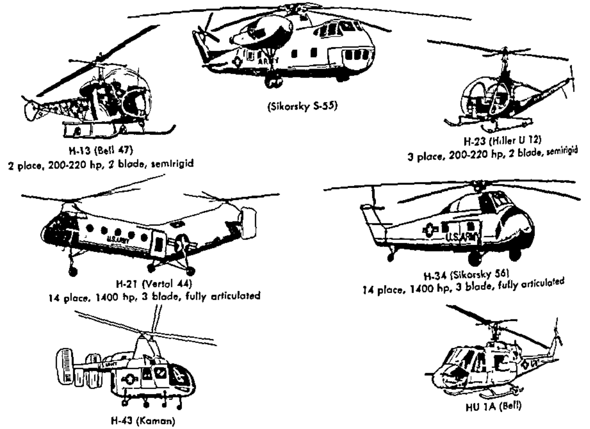


Fig. 7. Some typical helicopter types.

tary uses by most helicopter manufacturers. Aerodynamic and electronic methods of integrating and simplifying flight controls have been developed with the growing requirement for instrument flight rule operation of helicopters. Extensive experimentation in this area has been conducted by the U.S. Army, primarily at the Signal Corps testing activity at Fort Huachuca, New Mexico. One particular problem has been the absence of instruments specifically designed for the helicopter; normal aircraft instruments which primarily indicate the attitude, velocity, and other functions as a representation of the longitudinal axis of the fuselage itself are not applicable to a helicopter which may be flying sideward or backward from the axis of the fuselage. An Army-Navy program for the development of instruments portraying to the helicopter pilot his action with respect to the ground has been conducted, and numerous advances are being made.

Characteristics. The main measures of the helicopter are equivalent to those used in defining fixed-wing aircraft. The rotor geometry is described by the disk loading, which expresses the design gross weight as a function of the swept areas of the lifting rotor (or rotors), and by the solidity ratio, which defines the percentage of the disk area occupied by the rotor blades. These can be compared to the wing loading and aspect ratios, respectively, of fixed-wing aircraft. The power-to-weight ratio (the pounds of gross weight per horsepower), and the empty-to-gross-weight ratios tend to define the general performance characteristics.

Industry. The helicopter industry in the United States was given great impetus by large-scale military orders at the time of the Korean campaign. The vehicles saw a great deal of service, primarily as aerial battlefield ambulances and emergency resupply vehicles (Fig. 7). The oldest helicopter company is the Sikorsky Division of United Aircraft, with factories in Bridgeport and Stratford, Connecticut, where they have manufactured primarily single main-rotor helicopters ranging in size from the original two-place aircraft of World War II to the 3-ton Army H-37 which is the largest production helicopter made in the United States to date. Bell Aircraft Corporation has formed a separate helicopter corporation with facilities at Fort Worth, Texas, where they have manufactured several thousand of the smaller two-place helicopters for all military services and commercial users and several hundred larger tandem-rotor helicopters for Navy antisubmarine warfare application. Vertol Aircraft Company, originally Piasecki Helicopter Corporation in Morton, Pennsylvania, has manufactured large numbers of tandem-rotor helicopters for the Army, Navy, and Air Force, ranging in size from the five-place Army H-25 (Navy HUP) to the 21-passenger H-21 and the 30-passenger HC-1A. Kaman Aircraft of Windsor, Connecticut, has constructed a great number of dual lifting rotor Syncropters for the Marines and Air Force, and

is currently designing a single main-rotor helicopter for Air Force application. Hiller Helicopters of Palo Alto, California, has concentrated on a smaller, two-place, tilting-rotor helicopter, primarily for the Army and some commercial applicators.

In general, helicopter manufacturers have designed their own aircraft, and no independent helicopter design staffs have produced aircraft in numbers.

Three certified public helicopter carriers are located in California, Chicago, and New York. They have transported passengers and mail for a number of years in these metropolitan areas. Independent operators have been successful in the use of helicopters for agricultural spraying, fertilizing, and seeding, and for service to oil companies for pipeline surveys and transportation to the offshore Texas Towers. Many independent operators serve as local taxi operators and sightseeing services; they operate in cities such as Cleveland, St. Louis, and Washington.

Of the 5000–6000 helicopters produced to date, well over two-thirds are in service with the military departments, primarily with the Army for short-haul transportation in conjunction with military ground troops and as aerial ambulances and reconnaissance vehicles; with the Marine Corps as a means of transportation in amphibious assaults; and with the Air Force primarily for air-base rescue services. The Navy makes extensive use of the helicopter as a rescue vehicle with aircraft carriers (and has made some spectacular rescues using this combination), and as an antisubmarine vehicle carrying either sonar listening devices or weapons. Recent experiments conducted by the Army have shown the effectiveness of the helicopter as a close-in air-borne weapons carrier, an indication of a growing role for the helicopter in this area as well. The Navy has conducted tests with the helicopter in a towing role, where it has shown a tremendous ability in applying a tow force many times its own weight; a 10-passenger helicopter towed a destroyer at over 10 knots in one demonstration.

Professionally, helicopter activity is represented by the American Helicopter Society, which is affiliated with the Institute of the Aeronautical Sciences in New York, and the American Helicopter Association, an association of commercial helicopter operators, with headquarters in California.

Helicopters are designed and built in many other countries, and some are superior to American models. In England, primary helicopter activity is in the Westland Aircraft, Limited, which has used English versions of Sikorsky-designed aircraft with British power plants, and the Bristol Corporation, which has specialized, of late, in larger aircraft bordering on the convertiplane rather than the true helicopter. The most significant new British development is the Fairey Rotodyne, a gyrodyne or pressure-jet-powered helicopter with modified wings using aircraft-type propellers for increased forward velocity. France has produced a number of helicopter types. The Berguet Corporation is carrying on an organization originated by one of the earliest producers of successful helicopters and has primarily concentrated in the smaller aircraft area. Germany has produced a number of experimental designs. Both Vertol and Sikorsky have licensees constructing helicopters of various designs in these countries.

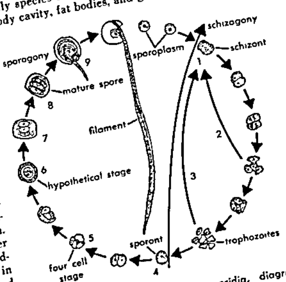
The Russians have shown considerable interest in the helicopter, primarily as a tactical vehicle for use by ground forces, although their extensive land areas and limited ground transportation would lead to the expectation that a considerable increase in commercial use of helicopters may be expected. Their helicopters are large and incorporate such features as blade anti-icing. The two principal helicopters with the Soviet military forces are similar in general appearance to the Sikorsky-type machines, being large, single main lifting rotor vehicles with antitorque tail rotors. A larger tandem-rotor helicopter has also been demonstrated.

[W.B.BN.]

Bibliography: A. Gessow and G. C. Meyers, *Aerodynamics of the Helicopter*, A. A. Nikolsky, *Helicopter Analysis*, 1951; P. R. Payne, *Helicopter Dynamics and Aerodynamics*, 1959.

Helicosporidia

An order of Cnidosporidia characterized by production of spores with a relatively thick, single intrasporal filament and three uninucleate sporoplasms. The spore membrane occurs as a single piece. *Helicosporidium parasiticum* Keilin is the only species in the order. The parasite infests the body cavity, fat bodies, and ganglia of mites (*Heri-*



Probable development of Helicosporidia, diagrammatic: 1–3, schizogony, with formation of the schizont; 4–9, sporogony, with development of mature spore; 4–7, sporont arising from trophozoites and development into immature spore; 7, mature spore; 8, open of mature spore; 9, escape of sporoplasms.

cia hericia), and the body cavity of fly larvae (*Dasyhelea obscura* and *Mycetobia pallipes*) found in the sap of horsechestnut and elm trees.

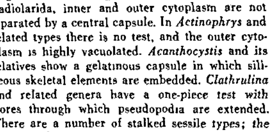
Although the complete life cycle of helicosporidians is not known, stages in the development and division of the trophozoite and in the development of the spore have been described. The sporont or sporoblast gives rise to 4 cells, 1 forms the spore membrane and 3 remain as sporoplasms. The origin of the filament is not known. *See* CNIDOSPORIDIA; *see also* DIPTERA; SARCOPTIFORMES. [R.F.N.]

Heliozoida

An order of the class Actinopodea frequently referred to as the Heliozoa. These protozoans are mostly fresh-water types and typically have radially arranged axopodia. In a few genera, such as *Acanthocystis*, the axonemes of the axopodia converge in a central granule. In contrast to the Radiolarida, inner and outer cytoplasm are not separated by a central capsule. In *Actinophrys* and related types there is no test, and the outer cytoplasm is highly vacuolated. *Acanthocystis* and its relatives show a gelatinous capsule in which siliceous skeletal elements are embedded. *Clathrulina* and related genera have a one-piece test with pores through which pseudopodia are extended. There are a number of stalked sessile types; the rest are floating organisms. *See* ACTINOPODEA. [R.P.H.]

Helium

Chemical element number 2, helium, He, is a member of the family of noble gases, group 0 of the periodic table. Helium is the second lightest element. The world's chief source of helium is a group of natural gas wells in the United States. *See* INERT GASES.



Uses. At one time, the only large use for helium was the filling of lighter-than-air balloons and dirigibles. The lifting power of helium is 92% that of hydrogen, and helium has the great advantage of being nonflammable. Relatively little helium is [illegible] phere to obtain cosmic-ray data.

The principal uses for helium are (1) in inert gas-shielded arc welding, (2) as a refrigerant (liquid helium) for obtaining and maintaining the lowest possible temperatures (0 to 5°K), and (3) for varied purposes in the atomic energy program and in the national defense. Because of the great demand for helium in connection with the second and third uses, argon is replacing helium, as far as possible, in the first use.

Helium has some physiological uses. Helium-oxygen mixtures are supplied instead of air to divers and caisson workers to prevent the painful disability known as the bends. Helium-oxygen mixtures are used in oxygen therapy for asthma patients whose breathing passages are constricted; helium flows much more easily than nitrogen, thus permitting the patient to receive more oxygen, diluted to a safe breathing level, per minute. Helium is sometimes used as an inert diluent for oxygen during anesthesia, but in most cases nitrogen is satisfactory for this purpose.

Other uses for helium are (1) as a gaseous cooling medium for nuclear reactors, since it does not become radioactive; (2) in wind tunnels to obtain design data on aircraft in simulated flights at extremely high speeds; (3) as an inert atmosphere in the production of titanium and zirconium metals and in the growth of silicon and germanium crystals for transistors (helium is being replaced, as far as possible, by argon in this application); (4) in mass spectrometer-type, highly sensitive leak detectors; (5) in filling neutron counters, gas thermometers, and x-ray spectrographs used to analyze materials containing light elements; (6) as a liquid in bubble chambers for detecting very high energy particles; (7) as the mobile phase in gas chromatography, one of the most important analytical tools ever discovered; (8) in gas-lubricated bearings; and also in various minor applications.

The demand for helium greatly exceeds the supply. An official release of the U.S. Department of the Interior stated that 90% of the helium produced in 1958 was allocated to agencies of and contractors for the Federal government, leaving only 10% for ordinary industrial and "civilian" uses.

Occurrence and origin. Helium constitutes 5.24 parts per million by volume of dry air. This helium is almost entirely the helium-4 isotope, and contains only 0.00013% by volume of helium-3. Helium-3 is 8 times more prevalent in helium taken from air than in helium from natural gas, and is 10 times more prevalent in helium taken from certain lithium metals than in helium from air.

Helium occurs in concentrations up to 2%, and occasionally higher, in certain natural gas wells in Kansas, Oklahoma, New Mexico, Texas, and Utah. Of the world's known recoverable helium, 99% is in natural gas fields located within 250 miles of Amarillo, Texas. Recently, the discovery of helium in natural gas in Russia and in the Union of South Africa has been announced.

Although it is generally supposed that the helium in natural gas originated as the result of radioactivity in the earth, there are still some unanswered questions. For instance: with uranium and other radioactive minerals widely scattered in the earth's crust, why is it only in certain isolated locations that appreciable concentrations of helium are found in natural gas?

Helium is occluded in appreciable concentrations in a few radioactive minerals. The highest known concentration (10.5 ml/g mineral) is found in thorianite. Traces of helium are widely scattered in many other minerals and in meteorites. A little helium is usually present in the gases in mines, volcanoes, fumaroles, and hot springs.

Helium is far more abundant outside the earth than on the earth; the best estimates are that only 0.000001% by weight of the earth's crust, including the atmosphere, is helium, whereas about 23% by weight of the visible universe (stars, nebulae, interstellar space) is helium.

Discovery. The existence of a bright yellow line in the spectrum of the sun's prominences was discovered during a solar eclipse on August 18, 1868, by six different observers at different places on the earth. On October 20, 1868, another observer, J. N Lockyer, observed three yellow emission lines in the sun's chromosphere, and recognized that one of these did not correspond to any of the known dark absorption lines in the sun's spectrum. In 1869, G. Rayet first definitely ascribed the new line to some element other than hydrogen or sodium. In 1871, Lockyer definitely recognized that the yellow line corresponded to a new element, which he and E. Frankland called helium. In 1895, Sir William Ramsay, in England, examined the spectrum of a gas liberated from a Norwegian uranium-thorium-lead mineral called cleveite, and found it to contain a brilliant yellow line which W. Crookes identified as the helium line. Thus, helium was found for the first time on earth.

Properties. Helium is a colorless, odorless, and tasteless gas. It has the lowest solubility in water of any known gas. It can be liquefied, but it has the lowest condensation temperature of any known substance. Under its own vapor pressure, helium is a liquid even very near the absolute zero. Helium can be frozen by the application of sufficient pressure at a temperature of 1.1°K or below. The lowest pressure at which solid helium can exist is 25 atm.

Of great interest to those who work in the field of cryogenics is the fact that helium forms two different liquid phases with distinctive properties. The liquid that exists above the transition temperature is called liquid helium I, and the liquid that exists below this temperature is liquid helium II. The transition temperature at the equilibrium vapor pressure of helium is called the λ-point, and is 2.19°K. Although helium I is a normal liquid, helium II is unique. Helium II is one of the best heat conductors known and has probably the lowest viscosity of any known liquid. It forms a thin film over any solid with which it comes in contact, and the liquid creeps through this film to other containers.

Properties of helium	
Atomic number	2
Atomic weight	4.003
Melting point,* at 25.2 atm pressure	−272.1°C (1.1°K)
Triple point (solid, helium I, helium II)	−271.37°C (1.78°K)
Triple point = λ-point (helium gas, helium I and helium II)	−270.96°C (2.19°K)
Boiling point at 1 atm pressure	−268.9°C (4.2°K)
Gas density at 0°C and 1 atm pressure, g/liter	0.17846
Liquid density at its boiling point, g/ml	0.122
Solubility in water, cm³ gas/1000 g water at 25°C and 1 atm pressure	9.26

* The melting point varies with the pressure.

See HELIUM, LIQUID.

Helium forms no chemical compounds in the ordinary sense of the word, although there is some evidence that weakly bonded ions with more than one atom can be formed under certain circumstances. The helium molecule is monatomic.

Production and distribution. Small quantities of helium are produced from monazite and other minerals, and from the air, but the only substantial source of helium is natural gas wells. Plants for separating helium from natural gas are operated for the Federal government by the Bureau of Mines and are located in Texas, Kansas, and New Mexico.

A release from the U.S. Department of the Interior Information Service states: "In 1957, the Bureau produced 310,000,000 cubic feet—five times the production in 1950. But, at the present time, 4,000,000,000 cubic feet are being wasted annually."

Helium is separated from natural gas in three main steps: (1) removal of carbon dioxide and water vapor, (2) liquefaction of almost all constituents present in the gas except the helium, and (3) purification of the partly refined helium by adsorption of the impurities on activated charcoal cooled by liquid nitrogen. The entire process is continuous.

In separating helium from air, all the constituents of the air are liquefied except helium, neon, and hydrogen. This crude uncondensed mixture is contaminated by nitrogen, most of which is removed by adsorption on silica gel at −190°C. The hydrogen is burned out, and the neon is removed by adsorption on activated carbon at a very low temperature.

A giant air-separation plant processing 5000 tons/day of air can produce no more than 560 ft³/day of helium, as compared to almost 1,000,000 ft³/day produced from natural gas in 1957.

In countries where there are no abundant natural sources of helium, it may pay to recover the helium liberated when certain radioactive minerals are processed. Helium has been recovered monazite sand containing between 0.5 an

occluded helium per gram. The quantities of thorianite (or of other minerals richer in helium than monazite) processed are so small that installation of helium recovery equipment is not practical.

Helium, unlike nitrogen, oxygen, and argon, has never been commercially transported as a liquid. For the larger uses, it is shipped under pressure in large metal cylinders, and for smaller uses, in standard compressed-gas cylinders.

Analytical determination. The principal methods of detecting the presence of helium are mass spectrometry, gas chromatography, and emission spectroscopy. The first two of these methods are also used for the quantitative determination of helium. A third quantitative method involves adsorbing all gases present in a sample except the helium on activated charcoal cooled by liquid nitrogen, measuring the volume or pressure of the residue, and then checking the purity of the residue by its emission spectrum. The only gases that are likely to interfere with this method are hydrogen and neon. If hydrogen is present, it can be removed over hot copper oxide. If neon is present, a second adsorption step on freshly activated charcoal at the temperature of liquid nitrogen will probably remove it, leaving only helium in the gaseous state.

[C.A.C.]

Bibliography: C. C. Anderson and H. H. Hinson, *Helium-bearing Natural Gases of the United States, Analyses and Analytical Methods*, U.S. Bur. Mines, Bull. 486, 1951; W. H. Keesom, *Helium*, 1942; ... and M. A. Whiteley (eds.), *Thorpe's Dictionary of Applied Chemistry*, 4th ed., vol. 6, 1943; H. P. Wheeler, Jr. and L. B. Swenarton, *Helium, Bibliography of Technical and Scientific Literature from its Discovery (1868) to Jan. 1, 1947*, U.S. Bur. Mines, Bull. 484, 1952.

Helium, liquid

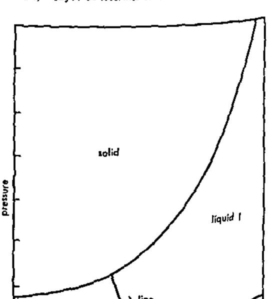
The unique properties of liquid and solid helium have led to their being considered by many as a distinct element. See HELIUM.

Liquid helium. The very low boiling point of helium (−268.9°C or 4.2°K) makes it of especial importance in research at temperatures near absolute zero, because liquid helium is used as a cooling medium ... temperature. For discussions of the methods of preparing liquid helium, see CRYOGENICS; LIQUEFACTION OF GASES.

By boiling under a reduced pressure (less than 0.01 mm), a temperature of 0.7°K has been obtained. Liquid helium-4 exists as two forms, helium I and helium II. Between these two forms, transition occurs along the λ line (2.178°K at 1 atm). Helium I is a normal liquid, but the properties of helium II are so unusual that it has been described as a fourth state of matter. If the properties of liquid helium I are observed as the temperature is lowered, there are seen discontinuous changes in the specific heat, viscosity, thermal conductivity, compressibility, and surface tension as the λ line is crossed. Molecular structure and molecular refractivity do not change at the λ line, but the temperature coefficients of density, vapor pressure, dielectric constant, and refractive index do change. The behavior of liquid helium resembles that of a gas, a so-called degenerate gas. The viscosity of helium II is very low (about 0.001 that of hydrogen gas), and the thermal conductivity is very high (about 800 times that of copper at room temperature). See SUPERFLUIDITY.

Although liquid helium I is somewhat gaslike because of its high zero-point energy, it has the properties of a liquid or of a reasonable classical aggregate; but liquid helium II is quite different. Liquid helium I boils vigorously if it is cooled by withdrawing vapor from the space above it. However, if the liquid is cooled to the λ point, the boiling stops abruptly. The difference in the behavior of the two liquids is explainable as resulting from an increase in thermal conductivity by a factor of about 1,000,000 accompanying the change from liquid I to liquid II at the λ point on the phase diagram.

All substances in contact with liquid II are covered by a film about 10^{-6} cm thick. If an empty glass beaker is partly immersed in a bath of liquid helium, a layer of helium can be seen to rise slowly



Phase diagram for helium-4 at low temperatures. (Not to scale.)

inside the beaker until it reaches the same level as the helium on the outside of the beaker. If the beaker is raised somewhat, the level of the liquid inside the beaker gradually falls until it is the same as that on the outside. If the beaker is completely removed from the bath, the level of the helium within the beaker continues to fall until it is empty. Small drops of liquid helium appear on the bottom of the beaker and fall into the bath. This phenomenon has been called film flow.

One striking property of liquid helium II is its tendency to flow from a region of low temperature to a region of high temperature if the two reservoirs are connected by a capillary. Conversely, if liquid II is caused to flow from one reservoir to another through a capillary, the liquid below the rising column will increase in temperature. The first of these is called the thermomechanical effect and the second the mechanocaloric effect.

Second sound is a temperature wave observed in liquid helium below the λ point, whereas ordinary sound (first sound) is a pressure wave. In terms of the two-fluid model of liquid helium, second sound represents a periodic oscillation of the concentration of the two fluids. At any given point in the liquid at an instant of time, the superfluid (helium II) is moving in one direction whereas the normal fluid moves in the opposite direction, resulting in no change in density. Because the superfluid component of liquid helium represents the cold or condensed state (zero entropy) and the normal fluid represents the warm or excited state, an increase in the concentration of the superfluid component at a point lowers the temperature, whereas an increase in the concentration of the normal component raises the temperature. Second sound may be considered as either a temperature wave or an entropy wave. The velocity of second sound is zero at the λ point of helium, rises to a maximum value of 20 m/sec at 1.6°K, passes through a minimum around 1°K, and then rises to a maximum of about 200 m/sec as the absolute zero is approached. Second sound was predicted independently by L. Tiza (1938) and L. Landau (1941) and first observed experimentally by V. Peshkov (1944).

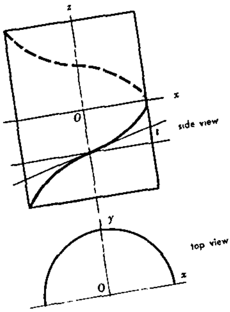
Liquid helium-3 does not exhibit phenomena similar to the two liquid forms of helium-4, but resembles a classical low-boiling liquid. The unique properties of liquid helium-4 can be used as bases for methods of separating the two isotopes of helium.

Solid helium. A minimum of 25 atm pressure is required to form solid helium, probably because of the low van der Waals forces associated with the electronic structure of the helium atom and because of the high zero-point energy. *See* LOW-TEMPERATURE PHYSICS.

Bibliography: C. J. Gorter (ed.), *Progress in Low Temperature Physics*, vols. 1 and 2, 1957; K. Mendelssohn, Superfluids, *Science*, 127(3239): 215–221, 1958.

Helix

Any nonplanar curve all of whose tangents make the same angle with a fixed line. Other characteristic properties are that (1) all principal normals are parallel to a plane, and (2) the ratio of torsion to curvature is constant (*see* GEOMETRY, DIFFERENTIAL). If a helix has constant curvature (and hence constant torsion), it is a circular helix; it lies on a circular cylinder whose elements it cuts at a constant angle (see illustration). Parametric equations for a circular helix (the curve of an untapered screw) are $x = a \cos t$, $y = a \sin t$, $z = kt$, $a > 0$, $k \neq 0$, and $-\infty < t < \infty$. *See* ANALYTIC GEOMETRY; PARAMETRIC EQUATION. [L.M.BL.]



Circular helix.

Hellbender

A large aquatic salamander, *Cryptobranchus alleganiensis*, found in certain of the larger streams and rivers of the eastern United States, most commonly in the Ohio River system. This is the largest of all American salamanders. Adult females reach a length of 27 in.; males are somewhat smaller. The adults are stoutly built, soft-bodied animals, with broad, stubby legs. The hellbender is of variable dark shades, ranging from brick red to brownish black in ground color, and spotted with lighter or darker markings. Gills are lost when the larva is about 18 months old.

The hellbender is nocturnal, hiding under sheltering objects on the stream bottom during the day. It is carnivorous, eating whatever animals it can catch, usually crayfish. The common name is said to be derived from the frantic contortions of this animal when hooked by an angler. Its flesh is said to be palatable, but it is rarely eaten.

There is a similar species, *Cryptobranchus bishopi*, in the streams of southeastern Missouri and northeastern Arkansas. *See* SALAMANDER; SALAMANDROIDEA. [J.D.B.]

Hellgrammite

The larva of the dobson fly, *Corydalus cornuta*, of the insect order Neuroptera. Flat, strongly segmented, fringed with gills, and armed with heavy mandibles, the hellgrammite is a fearsome-looking insect. When fully developed the larva is about 3 in. long.

Hellgrammites live as aquatic larvae for about 3 years. They are predaceous, preying primarily upon other insects. They are prized as fish bait, especially for bass.

The adults, with a wingspread of about 5 in., are among the larger insects of the United States. They have two pairs of large, membranous wings which are held flat over their back at rest. Eggs are laid in large masses over the water, and the larvae drop into the water as they hatch. Other members of the family Corydalidae, the willow flies and alder flies, are much smaller. Their larvae are also sometimes called hellgrammites, and are used for bait by trout fishermen. *See* NEUROPTERA. [J.D.B.]

Helmholtz coils

A pair of flat circular coils with equal numbers of turns and equal diameters, arranged with a common axis, and connected in series to have a common current (Fig 1). The purpose of the arrangement is to obtain a magnetic field H_{AB} that is more nearly uniform than that of a single coil (Fig. 2) without the use of a long solenoid. *See* SOLENOID (ELECTRICAL).

The optimum arrangement is that in which the distance between the two coils is equal to the diameter of one of the coils, as shown in Fig 1. For this arrangement, the variation of the field strength near the center of the apparatus is a minimum, and therefore the field is nearly uniform near the center. The field is the sum of the fields produced there by the individual coils. [K.V.M.]

Bibliography: L. Page and N. I. Adams, *Principles of Electricity*, 1958.

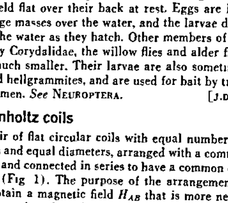


Fig 1. Arrangement of Helmholtz coils. (From L. B. Loeb, *Fundamentals of Electricity and Magnetism*, 3d ed., Wiley, 1947)

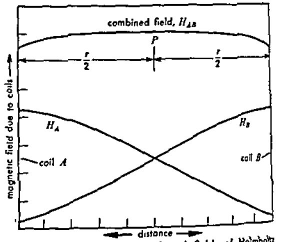


Fig. 2. Separate and combined fields of Helmholtz coils. Note the region of constant resultant field at point P. (From L. B. Loeb, *Fundamentals of Electricity and Magnetism*, 3d ed., Wiley, 1947)

Helobiales

An order of the plant subclass Monocotyledoneae including seven families with uncertain interfamily relationships. All are aquatic herbs, with mostly small flowers and seeds without endosperm. Some of the more familiar plants in this group are the pondweeds (*Potamogeton*), wapato, waterplantain, *Elodea*, and eel grass (*Vallisneria*). The Helobiales have a world-wide distribution but are of little economic importance. Members of the frog's-bit family (Hydrocharitaceae) are cultivated as aquarium plants. *See* EMBRYOPHYTA; MONOCOTYLEDONEAE; PLANT KINGDOM. [P.D.S.]

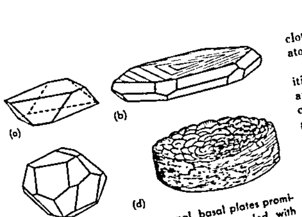
Hemadsorption viruses

Two viruses recently isolated from children with respiratory illnesses. These agents are detectable as yet only by means of the hemadsorption technique, a test involving adsorption of red blood cells to virus-infected cells in tissue culture. They have been classified with the parainfluenza viruses. It has been suggested that these new viruses might be among the etiological agents of the common cold. *See* ANIMAL VIRUS. [J.L.M.]

Bibliography: E. Jawetz, J. L. Melnick, and E. A. Adelberg, *Review of Medical Microbiology*, 3d ed., 1958.

Hematite

The most important ore of iron. As such, hematite is considered by many to contribute more than any other mineral to modern civilization. Its composition is Fe_2O_3, and when pure, it contains 70% iron. It crystallizes in the hexagonal system. Most commonly, crystals are basal plates, frequently bev-



Hematite crystals. (a-b) Hexagonal, basal plates prominent. (c) Hexagonal, edges of plates beveled with rhombohedral forms. (d) Thin plates grouped in rosette forms. (From C. S. Hurlbut, Jr., Dana's Manual of Mineralogy, 16th ed., Wiley, 1952)

eled by faces of a rhombohedron, on which there are often triangular markings. More rarely crystals are distinctly rhombohedral with nearly cubic angles. As kidney ore, it occurs in botryoidal or reniform shapes with radiating structure, and as specular iron ore it is micaceous and foliated (see illustration). It is most commonly fine grained and earthy. Hematite has no cleavage but some crystals show a basal parting as well as a well-developed rhombohedral parting at nearly cubic angles. For crystals the luster is metallic, the color is steel gray, the hardness is 5.5–6.5 (Mohs scale), and the specific gravity is 5.26. For earthy varieties the luster is dull, the color is red, and both hardness and specific gravity are less than for crystals.

Hematite is a widely distributed mineral. It is found in sedimentary rocks of all ages, as an accessory mineral in igneous rock, in contact metamorphic deposits, and in enormous masses in regionally metamorphosed rocks. In red sandstones, it is the cementing material that binds the quartz grains together. Most of the hematite that forms the great iron ores of the world is of the columnar and earthy varieties; more rarely it is specular or coarsely crystalline. Large deposits of hematite are found on all the continents and are thus mined in many countries. Among the large producing countries are Canada, England, Australia, Brazil, India, Russia, Morocco, the Union of South Africa, and Spain. The United States produces 40% of the world's iron ore, mostly as the mineral hematite. The major production since the latter part of the nineteenth century has been from the Precambrian rocks of the Great Lakes region. The oolitic hematite ores of the Clinton formation near Birmingham, Alabama, have been an important source. See ILMENITE; IRON (EXTRACTION FROM ORE); ORE AND MINERAL DEPOSITS; REDBED. [C.S.HU.]

Hematologic disorders

Those disorders marked by aberrations in structure or function of the blood cells or the blood clotting mechanism. Although many other diseases may be reflected by the blood and its constituents, the abnormalities of red cells, white cells, platelets, and clotting factors are considered to be primary hematologic disorders.

Erythrocytic abnormalities. Red-cell abnormalities are principally represented by the anemias and polycythemias. The anemias are marked by decreases in red-cell numbers or amounts of hemoglobin contained in them. They are considered separately. See ANEMIA.

Polycythemias are disorders characterized by an increase in the numbers of circulating red cells and usually by a concomitant increase in hemoglobin. Relative polycythemia is really the result of hemoconcentration in which there is fluid loss from the blood with corresponding increases in proportions of cellular elements. It is seen in excessive vomiting, diarrhea, dehydration, and similar conditions and represents no serious problem in itself.

Secondary polycythemias result from a compensatory increase in the formation of red cells following hypoxia of the bone marrow. This condition is most often associated with chronic conditions, such as heart or lung diseases, chronic carbon monoxide poisoning, and prolonged exposure to high altitudes. The blood volume may be considerably expanded and the patient may present a typically dusky-red appearance, congestion of capillary beds, and frequency of bleeding episodes from the nose, stomach, and elsewhere.

Polycythemia vera is an infrequent disease of older persons in which there is a gradual increase in the numbers of red cells in the circulation. The exact causes are unknown, although the disease has been likened to leukemia; other hematologists believe it results from an obscure form of bone marrow anoxia. See LEUKEMIA.

Leukocytic abnormalities. The many forms of white cell present in the circulation, the bone marrow, and the lymphoid tissues of the body give rise to a wide variety of conditions in which one or more of these elements appear in altered numbers or form.

The suffixes "-philia" and "-penia" denote increases and decreases, respectively, for the cells named. Thus there are leukopenia, neutrophilia, eosinophilia, and pancytopenia as examples of the wide range of possibilities. The absolute or relative increases or decreases of one type of leukocyte or all leukocytes are often characteristic of certain disease states, notably those due to infections. In addition to alterations in numbers, there may also be changes in the proportions of different cells present, or the appearance of certain immature or atypical forms.

The leukemias represent a special kind of malignancy in which there is usually an uncontrolled growth of one or more types of leukocytes, often reflected by tremendous increases in the white-cell count of the peripheral blood.

The leukemias are generally classified, on clinical grounds, as acute, subacute, or chronic. In addition, they are named to indicate the particular

cell which shows neoplastic characteristics. The most common forms of leukemia are lymphocytic and myelogenous leukemia. In the former, large numbers of abnormal or immature lymphocytes may be present in the blood; in the latter, some cell common to bone marrow is involved. Other varieties such as monocytic, eosinophilic, and plasma cell leukemia exist, but they are much less frequently encountered. *See* NEOPLASIA.

Many similarities exist in the clinical course and pattern of lesions seen in the leukemias. In the acute forms, the onset is abrupt, with the appearance of fever, malaise, and weakness. The course is usually rapid, with susceptibility to infections and a tendency toward hemorrhage accounting for many complications or leading to death.

Chronic leukemias may not be apparent for some time; quite often a patient is found to have leukemia during examination for the possibility of anemia or any other disease which produces weight loss, weakness, and other symptoms.

Hemorrhagic disorders. The hemorrhagic disorders result from a large number of known and unknown causes or contributing factors, often of a diverse nature. Basically, two mechanisms are involved, the susceptibility of blood vessels to injury, and defects in the clotting process. *See* HEMORRHAGE.

Blood vessel damage may occur as a result of direct or indirect damage by microorganisms during infections, as the result of vitamin C deficiency (scurvy), and following hereditary defects in blood vessel development. A number of comparatively rare disorders, usually with a hypersensitivity component, also fall into this category.

Defects in the clotting process may be due to hereditable factors, as in hemophilia, or to abnormalities of thrombin and fibrinogen formation, either congenital or acquired. Examples include congenital hypoprothrombinemia, a condition which follows liver disease.

Various forms of thrombocytopenia occur, all of which are characterized by a decrease of thrombocytes, or platelets, in the circulation. Since the platelets are of prime importance in the clotting process, such deficiencies result in increased bleeding tendencies.

In certain diseases marked by bleeding tendencies such as [illegible]

Hematopoiesis

The process of formation of the cellular elements of the blood. This is one of the most active and important processes occurring in the body. The magnitude of the activity of the hematopoietic system is immense due to the continuing need for replacement of the blood cells destroyed each day. Since the average life span of a red blood cell is only 120 days, approximately 250,000,000,000 erythrocytes must be produced daily to replace those which are destroyed. Serious consequences, vital to the well-being of the body, result if this system is damaged by disease, genetic lesion such as sickle-cell anemia, or through irradiation with x-rays or other high-energy radiations. In fact, the lymphocytes are more sensitive to ionizing radiations than nearly any other cell type in the body. Changes in the lymphocytes, therefore, are a good index of radiation damage.

SITE OF ORIGIN OF BLOOD CELLS

The reticuloendothelial tissue is characterized as a loose, fibrous, highly vascularized mesh, or stroma, of silver-staining fibers and fixed endothelial and macrophage cells. Within the spaces of this stroma are found the free "blast" cells which will give rise to the definitive blood cells. The [illegible]

phocytes, macrophages, and monocytes of the blood and is localized primarily in the lymph nodes of the lymphatic system and also in the spleen, thymus, and to some extent in the bone marrow.

Myeloid tissue. This tissue is normally limited to the red marrow of the ribs, sternum, vertebrae, and the proximal ends of the long bones of the body and is concerned with the production of the red cells (erythrocytes), granular leukocytes (eosinophiles, basophiles, and neutrophiles), and the megakaryocytes.

It must be emphasized that this division into lymphoid and myeloid tissues is one of convenience only. During embryonic development and in certain blood disorders this separation breaks down. For example, granular leukocytes may develop in such lymphoid tissues as the thymus, spleen, and lymph nodes in extramedullary myelopoiesis.

There are several theories concerning blood cell formation. They differ mainly in the number of primary "blast" cells which are invoked as precursors of the definitive cells. The chief controversy centers around whether the tissue and circulating lymphocytes are identical with the hemocytoblasts, the mother cells, of the myeloid tissue and hence are capable of giving rise to other, different cell types under the proper conditions, or whether they are themselves the end product of a process of differentiation and are capable only of forming other lymphocytes. The unitarian view holds that there is one basic pluripotential cell, indistinguishable from and identical with a lymphocyte, which gives rise to all other blood cells. This cell is called a hemocytoblast. The hemocytoblast is a large cell, up to 15 μ in diameter, whose cytoplasm shows an affinity for basic dyes. It is found in both lymphoid and myeloid tissues and is derived, in some unknown fashion, from one of the fixed reticular cells of the stroma of the reticuloendothelial system. This is in turn derived from the mesenchymal loose connective tissue of the embryo. Graded intermediate forms link the hemocytoblast with the production of lymphocytes, macrophages,

megakaryocytes, and, perhaps, the monocytes in a fairly direct pathway to each of these specialized cells. Less direct but more fully substantiated pathways lead to the differentiation of the erythrocytes and the granular leukocytes. The hemocytoblast gives rise to erythroblasts and granuloblasts, or myeloblasts, which yield the definitive mature cells. The plate shows the erythrocytic, granulocytic, and lymphocytic series as they appear in stained smears from bone marrow. It should be remembered that the cells have been flattened out into thin sheets in the smearing process and hence do not show the precise size and shape of the living cells.

Erythroblasts. These are spherical cells, slightly smaller than the hemocytoblast and with a more densely staining basophilic cytoplasm. As they develop toward erythrocytes, the cytoplasm becomes more acidophilic as hemoglobin accumulates, the nucleus shrinks and its chromatin becomes more dense until an intermediate form, the normoblast, is formed. The normoblast is characterized by bright pink cytoplasm, when stained with eosin, and a densely staining nucleus. The normoblast then loses its nucleus through extrusion to yield the definitive erythrocyte.

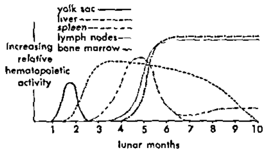
Granuloblasts. These cells, also known as myeloblasts, arise from the hemocytoblasts and show their specific granulation, as eosinophiles, basophiles, or neutrophiles, early in their development. Once their specific granulation has appeared, they have never been observed to transform into any other type of cell, so it is presumed that their developmental fate is fixed at that time. The maturation process involves an intensification of the staining of the granules, a condensation of the nucleus into a denser, horseshoe shape, and its final lobulation which is specific for the cell type it is to become. Many mitoses occur during this process, with the granules being evenly distributed at each division.

Megakaryocytes. These are thought to arise from hemocytoblasts by a process of repeated nuclear division followed by subsequent fusion of the nuclei without cytoplasmic cleavage. These giant cells ultimately degenerate, the fragmenting cytoplasm giving rise to the thrombocytes, or blood platelets, which participate in the clotting reaction.

Monocytes. These cells are believed to be derived directly from hemocytoblasts (or lymphocytes) in the venous sinuses of the bone marrow, liver, and spleen. This is, however, one of the most contested and imperfectly understood phases of hematology.

EMBRYONIC DEVELOPMENT OF BLOOD CELLS

The development of the blood cells occurs in a regularly ordered sequence of hematopoietic sites which is graphically illustrated for man. In the chart it will be seen that as development proceeds, one site after another shows the initiation of blood cell production. As a given organ exceeds and replaces the production of the preceding center, there



Graphic representation of hematopoietic activity in man during the normal gestation period of 10 lunar months of 28 days each.

is a decrease in activity of the organ as succeeding centers arise, until the adult pattern is established at about the time of birth. At his time the bone marrow and the lymphatic tissues predominate in myeloid and lymphoid element production, respectively. The course of these events will be described for each organ in man.

Areas of hematopoietic potency. From observations and experiments it has been determined that there are specific areas in which the blood cells first appear in the embryo. These areas of hematopoietic potency are essentially erythropoietic primordia since the first blood cells to arise in the embryo are the primitive erythrocytes. Within these areas, cells of the mesodermal component of the yolk sac coalesce to form stellate interconnected groups of highly basophilic cells (hemocytoblasts), called blood islands. These blood islands and their interconnecting branches become patent, as a result of the dissolution of the central portion of the cell mass and the secretory activity of the remaining epithelial cells to form the primitive plasma-filled vascular network. The vascular bed of the yolk sac serves a dual purpose in the lower vertebrates, being first a hematopoietic organ and later transporting foodstuffs from the yolk to the embryo. In mammals, where yolk is lacking, the yolk sac serves for hematopoiesis only [illegible] degenerates [illegible]

other cell types. Nearly all of the first formed hemocytoblasts give rise to primitive erythroblasts which rapidly form hemoglobin and become the primitive erythrocytes of early embryonic life. These primitive erythrocytes retain their nucleus and are slightly larger than the definitive series of erythrocytes which are enucleate (in mammals) and appear later in development. Later a few granular leukocytes are produced, but the hematopoietic activity of the yolk sac declines as the liver begins to function as the second major site of blood cell formation.

In the liver the free hemocytoblasts give rise to erythrocytes of the definitive enucleate type, ch[illegible]acteristic of late embryonic and adult life.

lar leukocytes and macrophages are produced also before the liver in turn begins to decline gradually in hematopoietic activity about the fifth month, with almost complete cessation of activity at birth.

The embryonic spleen produces mainly erythrocytes and granular leukocytes. Postnatally, this myeloid function is replaced by the formation of lymphoid cells, as indicated on the spleen curve showing the continuation of such activity in the illustration, and by the destruction of erythrocytes in the spleen.

The marrow, of the long bones especially, becomes the site of erythroblast and erythrocyte production in large numbers. The three types of granular leukocytes, megakaryocytes, and a few lymphocytes also arise in the marrow. See SKELETAL SYSTEM.

The lymph nodes and other lymphatic tissues arise rather late in development. In these tissues many large and small lymphocytes are formed from the free hemocytoblasts. These lymphocytes also undergo mitosis which increases their number appreciably. In addition, a few megakaryocytes and erythroblasts arise in the lymphatic tissues, along with more numerous granular leukocytes of the neutrophile and eosinophile types. These myeloid elements are no longer found in the lymphatic tissues after birth, except in certain diseases.

The thymus is an important site of lymphocyte production and, in the embryo only, of a few granular leukocytes as well. See THYMUS GLAND.

The hemoglobin [illegible]

that reason it has been termed fetal hemoglobin. The fetal form has a higher affinity for oxygen, is generally more resistant to destruction by alkali, and differs serologically from the adult form. Shortly after birth, fetal hemoglobin declines in amount and is gradually replaced by the adult form except in a few anemias where considerable amounts of the fetal hemoglobin may persist throughout life. See CARDIOVASCULAR SYSTEM; see also CIRCULATION DISORDERS.

[R.G.BD.; J.D.E.]

Bibliography: M. Bessis, *Cytology of the Blood and Blood-forming Organs*, 1956; H. Downey (ed.), *Handbook of Hematology*, 4 vols., 1938; A. A. Maximow and W. Bloom, *A Textbook of Histology*, 7th ed., 1957.

Hemicellulose

The term adopted by E. Schulze in 1891 to designate the plant-cell components which are made soluble by dilute alkali or which go into solution quite readily in hot dilute mineral acids with the formation of simple sugars. Today, the term is usually applied to those polysaccharides in the cell wall of land plants which are extractable by dilute alkaline solutions. The term has also been used to include all the polysaccharide components of the cell wall other than cellulose. See CELL WALLS IN PLANTS.

How the hemicelluloses, which are often water soluble after extraction from plant tissue, are anchored in the cell wall so as to be unextractable by water is still a controversial question. As an explanation, hemicellulose-lignin or hemicellulose-cellulose covalent bonds have been postulated. Hemicelluloses extracted from different plant sources, although they often have many common characteristics, are rarely identical. In fact, many different hemicelluloses usually occur intermixed with each molecular type representing different degrees of polymerization. Because of this heterogeneity, few hemicelluloses have been isolated in a homogeneous state. Thus, relatively little is known of the structure of these compounds that compose almost one-third of the carbohydrates in woody tissue.

Uses. Hemicelluloses are important to the paper industry. In chemical wood pulps, hemicellulose is needed for satisfactory pulp quality. Its presence aids the swelling of the pulp, the bonding of the fibers, the bursting strength, tensile strength, tear resistance, folding endurance, opacity, and specific surface of the pulp sheet. The optimum hemicellulose content of a pulp will vary, depending on the type of wood and the conditions of the pulping operations. Excessive amounts of hemicellulose will produce a brittle or glassine paper. See PAPER AND PAPER PRODUCTS.

On the other hand, if pulps are to be used for purposes other than papermaking, the presence of hemicellulose may be quite disadvantageous. For example, in the production of pulps which are to be used in the manufacture of rayon, cellophane, and cellulose esters and ethers, it is often essential to obtain an α-cellulose very low in hemicellulose. α-Cellulose is insoluble in 17.5% sodium hydroxide solution. In cellulose acetate manufacturing, small amounts of mannan or xylan cause hazy acetate solutions and clogged filter presses.

Hemicelluloses also serve as nutrients for yeasts, and they can be used for raw material in the production of furfural and ethyl alcohol.

Preparation. Hemicelluloses are generally prepared by alkaline extraction of dried, defatted plant tissue or a special delignified plant tissue which is called holocellulose. The latter is usually prepared by treatment of the defatted plant with chlorous acid solution, which solubilizes the lignin but leaves the polysaccharide behind in the same morphological relation as in the original plant tissue. Hemicelluloses are more readily extractable from holocellulose than from dried, defatted plant tissue, and can be divided into two solubility classes, hemicellulose A and hemicellulose B. Hemicellulose A is insoluble in slightly acidic solution, and therefore precipitates from the alkaline extract upon acidification. This class is composed mainly of high-molecular-weight linear glycans that contain small amounts of uronic acid. The degree of polymerization (DP) of these polysaccharides usually is 70–200. In comparison, the DP of cellulose is 2500–3500. For most plants, the A

fraction is the largest because within this fraction occurs the abundant and widely distributed polysaccharide, xylan. Hemicellulose B, which is precipitated from the neutralized alkaline extract by the addition of excess methanol or ethanol, consists of branched-chain molecules of a lower degree of polymerization than hemicellulose A. Its uronic acid content is higher than that of hemicellulose A. Further fractionation procedures, such as the formation of a complex between the hemicellulose molecules and copper salts and fractional precipitations with organic solvents, must be used to isolate a homogeneous polymer.

D-Xylose is the dominant building unit of the hemicelluloses of most woods and annual plants. The xylose content of hemicellulose from hardwood is generally higher than that from softwood. D-Mannose is also very abundant in hemicelluloses; the mannose content of softwoods is usually higher than that of hardwood. Often it occurs as a polymer, mannan, or in combination with D-glucose as a glucomannan. See CELLULOSE; MANNAN.

[J.L.SA.; R.L.WH.]

Hemichordata

A group of animals thought by some authorities to be a phylum of deuterostomes and by others to be a subphylum of the Chordata. At present, this group is considered to include the Enteropneusta, Pterobranchia, and Graptolithina. All are marine species found mostly in shallow waters, distributed from the Arctic to the tropics. These animals are superficially dissimilar and are believed to be derived from ancestors which had a stomocord (a short extension of the gut forward into the proboscis) as exemplified in living forms by the acorn worm Some zoologists have considered this structure to be homologous with the notocord of the Chordata, even though it is confined to the preoral region. Gill slits are present and the enterocoelic coelom is divided into three successive segments which are retained in the adult. A superficial nervous system is present and consists of cells and fibers, which may be slightly concentrated and hollow in a dorsal position. A tail and atrium are lacking in these species. The closest affinities of this group are probably with the Tunicata (Urochordata). See ANIMAL KINGDOM; CHORDATA; ENTEROPNEUSTA; GRAPTOLITHINA; PTEROBRANCHIA. [T.H.B.]

Bibliography: P. P. Grassé et al., *Echinodermes, Stomocordés, Procordés*, in P. P. Grassé (ed.), *Traité de zoologie*, vol. 11, 1948; C. J. Van der Horst, *Hemichordata*, in H. G. Bronn (ed.), *Klassen und Ordnungen der Tierreichs*, vol. 4, part IV, book 2, 1927–1935.

Hemicidaroida

An order of Echinacea with one very large tubercle on each interambulacral plate. In this respect they resemble cidarids, though the aboral tubercles may be reduced. The adoral ambulacral plates tend to be diademoid. There are four families. (1) The Orthopsidae, of the Jurassic and Cretaceous, had a camarodont lantern; the remaining families are stirodont. (2) The Hemicidaridae, of the Jurassic and Cretaceous, had the ambulacra abruptly widened at the ambitus and perforate, usually crenulate tubercles. (3) In the Acrosaleniidae, of the Jurassic and Cretaceous, the anus is displaced to one side of the periproct by some extra suranal plates. The tubercles are perforate and crenulate. (4) The Saleniidae, resembling the Acrosaleniidae, have imperforate tubercles. They range from the Jurassic onward, with two surviving deep-sea genera. See ARBACIOIDA; ECHINACEA; ECHINOIDEA.

[H.B.F.]

Hemidiscosa

An order of the subclass Amphidiscophora in the class Hexactinellida. These sponges are distinguished from the order Amphidiscosa in that the birotulates are hemidiscs with asymmetrical ends. *Hemidiscella* from the Cretaceous is an example. See AMPHIDISCOPHORA; HEXACTINELLIDA.

[W.D.H.]

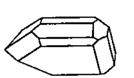
Hemimetabola

A division of the subclass Pterygota (winged insects). The orders included in this category are characterized by undergoing an incomplete metamorphosis during development. The immature, terrestrial, winged forms are referred to as nymphs, whereas the gilled, aquatic larvae are naiads. See PTERYGOTA. [C.B.C.]

Hemimorphite

A mineral sorosilicate with composition $Zn_4Si_2O_7(OH)_2 \cdot H_2O$; an ore of zinc. It crystallizes in the orthorhombic system, pyramidal class, and thus the prismatic crystals have different forms at top

is perfect prismatic cleavage. Botryoidal and staclactic aggregates, showing a crystalline surface and frequently impure, are common. Crystals are usually colorless and the aggregates white, but in some cases there are faint shades of green, yellow, and blue The mineral has a vitreous luster, a hardness of 4½–5 on Mohs scale, and a specific gravity of 3.45. Hemimorphite frequently resembles the zinc carbonate, or smithsonite, but can be distinguished by the reaction with hydrochloric acid. Hemimorphite is slowly soluble but smithsonite effervesces in cold acid. See PIEZOELECTRICITY; PYROELECTRICITY; SILICATE MINERALS.



Prismatic crystal of mineral hemimorphite, showing different forms at top and bottom. (From C. S. Hurlbut, Jr., Dana's Manual of Mineralogy, 16th ed., Wiley, 19

Hemimorphite is a secondary mineral found in the oxidized portion of zinc deposits associated with smithsonite, sphalerite, cerrussite, anglesite, galena, and more rarely the oxidized minerals of copper. It has a wide distribution and has been mined in Belgium, Germany, Rumania, England, Algeria, and Mexico. In the United States it is found at Sterling Hill, New Jersey; Freidensville, Pennsylvania; and Elkhorn Mountains, Montana. [C.S.HU.]

Hemiptera

An order of the class Insecta sometimes referred to as the Heteroptera. Both these names refer to the forewings which are differentiated into a thickened basal area and a membranous apical region. These are the true bugs. The related order, or suborder, Homoptera has forewings of uniform texture. Included in Hemiptera are such common insects as the bed bugs, stink bugs, plant bugs, lace bugs, and backswimmers. *See* HOMOPTERA.

The true bugs number about 25,000 species. They are known from all continents except Antarctica, and occur on most [illegible] w[illegible] oc[illegible] a[illegible] ground-inhabiting forms approximately 2 mm in length to the giant water bugs that are 100 mm or more.

Fig. 1. *Triatoma pallidipennis* (Stal). (Photograph by E. S. Ross)

In habits, the true bugs range from strictly phytophagous types attached to a single host plant to general predators on other insects and even to specialized ectoparasites of bats. Many species are of economic importance as plant pests or vectors of disease. They occur in vegetation, on the ground, in and on the water, and in the nests of termites. One family, the Aradidae, lives on fungi on the bark of dead trees, and certain Reduviidae and Cimicidae live in caves. Among aquatic forms, members of the genus *Rhagovelia* swim on riffles of swift-flowing streams, and *Aphelocheirus* remains permanently submerged in rivers. Most water bugs depend on surface air held to the body by air spaces and hairs on the abdomen. As oxygen is depleted in the air bubble through respiration, it is replaced from the surrounding medium by diffusion.

A few Hemiptera, like *Dacerla*, *Coquillettia*, and *Pilophorus*, mimic ants. Others, including cotton stainers (*Dysdercus*) and certain reduviid predators (*Phonoctonus*), resemble distasteful cantharids and other insects of orange or yellow and black colors which is a protective Müllerian type of mimicry. Still other bugs resemble their surroundings and hence are concealed from birds, lizards, and other enemies. One of the most striking of these is the flat, barklike *Phloeoa* of South America (*see* PROTECTIVE COLORATION). Stridulation, or sound production, is less conspicuous in Hemiptera than in the cicadas of Homoptera and katydids of Orthoptera but is developed in several groups. Stink bugs, flat bugs, the predaceous reduviid bugs, and several aquatic groups are examples. The sounds are produced either by rubbing filelike portions of the legs against knife edges on the body wall, rubbing parts of the wings against the body, or by rubbing the beak along a prosternal cross-striated stridulatory furrow, as in the Reduviidae (Fig. 1).

Most Hemiptera are bisexual and oviparous, but parthenogenesis is known and a pseudoplacental organ results in a special type of viviparity in Polyctenidae. Mating usually takes place on vegetation or on the ground, the pairing being end to-end in stink bugs, squash bugs, chinch bugs, and similar species, and with the male above the female in most others. Bedbugs (Cimicidae) have a special organ of Ribaga (or organ of Berlese) through which spermatozoa are introduced to the hemocoel of the female where they penetrate the ovarioles and fertilize the eggs. *See* INSECT PHYSIOLOGY; INVERTEBRATE EMBRYOLOGY.

Classification. The Hemiptera were first divided by P. Latreille, in 1825, into Hydrocorisae (water bugs) and Geocorisae (land bugs). L. Dufour, in 1833, proposed a third group, Amphibicorisae (surface water bugs). F. Fieber, in 1851, substituted the Cryptocerata (hidden antennae) for Hydrocorisae and Gymnocerata (exposed antennae) for Geocorisae. Both systems are in use today, but the Latreille-Dufour system conforms best to recent studies. In 1954, the Geocorisae were divided into Cimicomorpha and Pentatomomorpha by D. Leston, J. Pendergrast, and T. Southwood. The names of subfamilies and superfamilies have been in a state of flux since G. Kirkaldy's use of the "oldest genus principle."

The following table represents a balance between extremes and includes each distinctive group, down to the family level, as summarized by W. China in 1956 (Fig. 2).

Table 1. Families of Hemiptera

Classification	Common name	Distribution	No. of species
Suborder Hydrocorisae			
Corixidae	Water boatmen	General	300
Nepidae	Water scorpions	General	170
Belostomatidae	Giant water bugs	General	140
Notonectidae	Backswimmers	General	170
Pleidae	None	General	20
Helotrephidae	None	Tropical	20
Naucoridae	Creeping water bugs	General	200
Gelastocoridae	Toad bugs	Tropical and subtropical	80
Ochteridae	Velvety shore bugs	Tropical and subtropical	20
Suborder Amphibicorisae			
Gerridae	Water striders	General	300
Veliidae	Smaller water striders	General	200
Hydrometridae	Marsh treaders	General	50
Mesoveliidae	Water treaders	General	20
Hebridae	Velvet water bugs	General	40
Suborder Geocorisae			
Saldidae	Shore bugs	General	200
Leptopodidae	None	Tropical and subtropical	20
Leotichidae	None	Oriental	2
Aepophilidae	Marine bugs	Palearctic	1
Dipsocoridae	None	General	30
Schizopteridae	Jumping ground bugs	Tropical and subtropical	70
Cimicidae	Bat, bed, bird bugs	General	50
Anthocoridae	Flower bugs	General	300
Polyctenidae	Bat bugs	Tropical and subtropical	20
Miridae	Plant bugs	General	5000
Isometopidae	Jumping tree bugs	General	50
Microphysidae	None	Palearctic, tropical, oriental, Australian	30
Thaumastocoridae	Palm bugs	Neotropical	11
Joppeicidae	None	Mediterranean	1
Vianaidae	None	Neotropical	2
Nabidae	Damsel bugs	General	250
Enicocephalidae	Gnat bugs	General	100
Reduviidae	Assassin bugs	General	3000
Phymatidae	Ambush bugs	General	120
Tingidae	Lace bugs	General	700
Aradidae	Flat bugs	General	800
Termitaphididae	Termite bugs	Tropical	10
Lygaeidae	Lygaeid bugs	General	2000
Colobathristidae	None	Tropical	70
Berytidae	Stilt bugs	General	100
Piesmidae	Ash-gray leaf bugs	General but discontinuous	20
Pyrrhocoridae	Pyrrhocorid bugs	General	400
Coreidae	Coreid bugs	General	2000
Hyocephalidae	None	Australia	1
Pentatomidae	Stink bugs	General	2500
Phloeidae	Stink bugs	Neotropical Borneo	6
Plataspidae	None	Old World	400
Cydnidae	Ground or burrower bugs	General	600
Urostylidae	None	Asiatic and Australia	50
Aphylidae	None	Australia	2

MORPHOLOGY OF HEMIPTERA

External anatomy. In Hemiptera and Homoptera the mouthparts are elongate and slender, forming a sucking mechanism with sheathlike labium and needlelike mandibulary and maxillary stylets (Fig. 3). In Homoptera, as in leafhoppers and aphids, the beak arises from the posterior part of the head whereas in Hemiptera the position is anterior, and the head is commonly directed forward or downward rather than backward.

The stylets are grooved and slide up and down during the act of feeding, the mandibles being toothed to grip the plant or other tissue that is being penetrated. Two tubes are formed by the stylets, one for sucking up the fluid food and the other for injecting salivary or anticoagulant fluids. The labium is jointed with three or four segments.

Hemiptera are further characterized (Fig. 4) by antennae, usually of four or five segments, a pair of compound eyes, and often two ocelli. The thorax consists of a prominent pronotum, a triangular mesothoracic scutellum, and a broad metathorax which is partly fused with the first abdominal segment. The mesothoracic wings, or hemelytra, overlap at their membranous apices when at rest, leaving the scutellum exposed. They consist of a leathery corium at the base, an inner clavus along the scutellum, and an apical membrane. A marginal fracture separates the apex of the corium into a cuneus in some groups. The venation is seldom clear but has been interpreted from the tracheae of nymphal wing pads as consisting of the subcosta at the anterior edge of the wing, the radius and media more or less fused behind, the cubitus as the posterior vein of the corium, and one or more anal veins in the clavus. The hindwings are hitched to the forewings in flight by grooves and pegs. In metathoracic wings, the subcosta, radius, and media are variously fused but usually form an elongate cell with or without a backward-directed spur vein or hamus. The anal and jugal areas are more or less developed with ill-defined veins and folds.

Wings are sometimes reduced to short pads and may be lacking in certain groups or even in members of a single species. Wing polymorphism,

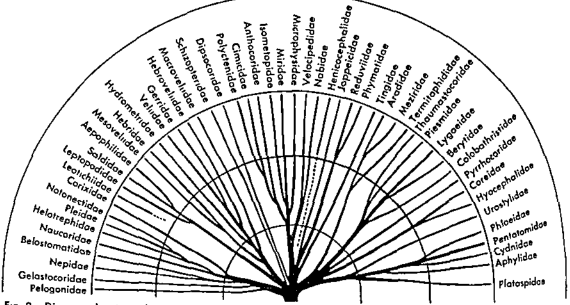


Fig. 2. Diagram showing relationships of families of Hemiptera. (*After W. China*)

pecially in water-striders, appears to be determined by environmental factors as well as by genetic factors.

The front legs are frequently enlarged and sometimes chelate in predaceous forms, and the middle and hind legs are adapted for swimming in some groups. The foretibiae have a small plate or comb at the inner apex and sometimes a spongy pad for

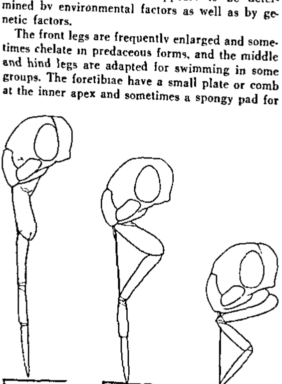


Fig 3. Mouthparts and method of feeding of a plant bug

clinging to smooth surfaces or holding prey. Tarsi are usually 3-segmented, but some groups have only two segments and front tarsi are lacking in a few specialized types. Two claws are commonly present with or without leaflike or bristlelike arolia between.

The abdomen is of 10–11 segments, but only [illegible] being fused [illegible] ns [illegible] gans are shown for both sexes in copulation (Fig. 5). Males have the ninth segment variously developed as a genital capsule. A pair of claspers is commonly associated with the genital capsule, and the aedeagus, or penis, arises from its floor. The aedeagus is hinged to basal plates. An ejaculatory duct traverses the entire structure, terminating in the gonopore. Clasping organs may be present on either side of the thickened phallosoma, and various lobes and processes are developed on the often membranous endosoma. The latter is evaginated during mating, with the vesica variously inflated and inserted in the female genitalia. The female abdomen is variously modified but commonly has an ovipositor, with bent valves, arising from the eighth and ninth segments. A spermatheca is present in many groups and is distinctive in shape and form in some. Spiracles occur laterally, either in a ventral position or dorsally and most often in or near lateral connexival plates. Long sensory hairs, or trichobothria, occur on the ventrites of stink bugs, squash bugs, chinch bugs, and some others.

Internal anatomy. Internally the Hemiptera resemble other orders of insects, but several special features are noteworthy. Parts of the midgut have fluted caeca in stink bugs, squash bugs, and some others. These contain symbiotic bacteria which are specific and are transmitted from one generation

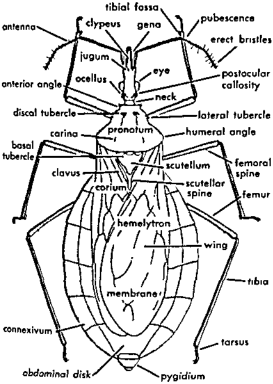


Fig. 4. External anatomy of a bug.

to the next, in or on the eggs. In stink bugs the eggs are smeared at the time of laying, and the first stage nymphs, when hatched, cluster around the egg shells and suck up bacterial material. Cimicidae have no gastric caeca but possess a pair of mycetomes in about the third abdominal segment. The function of symbiotic bacteria is not clear, but in the Reduviidae, R. Wigglesworth has shown that vitamins of the B group are produced.

Salivary glands are of two parts, the bilobed principal glands and the tubular accessory glands. They empty into a common duct and thence to a pump which forces the fluid through the stylet tube. The Malpighian tubules are usually four in number.

Respiration is accomplished by spiracles, of which there are 10 pairs or less. In terrestrial forms a closing apparatus is developed to conserve moisture, whereas in water bugs, the spiracles are open. Breathing tubes are present in Nepidae, enabling these water scorpions to use surface air while submerged. *Buenoa* and related backswimmers have hemoglobin in clusters of cells adjacent to the spiracles

Endocrine glands include the corpus cardiacum and corpus allatum located behind the brain. These *control molting and metamorphosis.* The presence of hormone from the corpus allatum (juvenile, or inhibitory, hormone) prevents metamorphosis in young nymphs Later, after four or five nymphal instars, the inhibitory hormone is no longer present and the growth and differentiation hormone acts to produce metamorphosis

Scent glands are a characteristic feature of bugs, causing the well-known "buggy odor." Most adult bugs have a pair of metathoracic glands with openings on the metapleura, as in Pentatomidae and others, or on the hind coxae, as in Reduviidae.

A few have additional glands in the first abdominal segment. Nymphs of all but a few, such as the Triatominae and Hydrometridae, have dorsal ab-

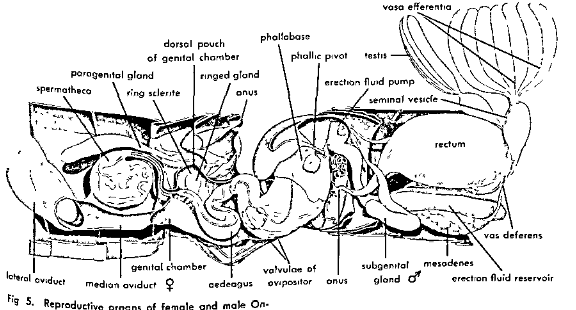


Fig 5. Reproductive organs of female and male *Oncopeltus fasciatus* (Dallas) in copulation. (*After P. Bonhag*)

the third and fourth or the fourth and fifth tergites; one, on the fourth tergite; to none. The openings may be single (Miridae) or double and widely separated (Naucoridae). In adults of the water striders (Gerridae), there is a single opening (omphalium) at the middle of the metasternum.

Internal reproductive organs include, in females, a pair of acrotrophic ovaries with ovarioles numbering 2–8, with 7 as the commonest number. A spermatheca is well developed and of characteristic shape for each of the main groups of stink bugs, squash bugs, and others, but is absent in plant bugs (Miridae) and their allies. In males, the paired testes comprise 1–7 follicles. The seminal vesicles are usually large and are formed by a glandular reservoir. Accessory glands known as mesadenia empty into the reservoir, and ectadenia join the ejaculatory bulb in some groups.

Chromosomes are best seen in late instar nymphs or during early adult life in males. The $2n$ (diploid) number may be as low as 6, in *Rhytidolomia*, or as high as 43, in *Ranatra*. Commonly, there are 12 autosomes and 2 X chromosomes or an X and a Y. In *Cimex* and some others there are supernumerary X chromosomes due to fragmentation (*see* GENETICS).

Development and metamorphosis. Hemiptera undergo simple or incomplete metamorphosis including egg, nymph, and adult stages (Fig. 6). Eggs are characterized by a chorion of several layers, usually with hexagonal reticulations on the surface. A lid or cap is present in the plant bugs, bedbugs, and allies, but the anterior end of the egg is marked only by micropylar processes in other groups. An egg burster is well developed and T-shaped in Pentatomidae and is less developed but visible in some other groups. Eggs are glued to surfaces by *Lethocerus*, inserted in plant tissues by Miridae, or laid free by some *Triatoma*. In *Belostoma* and *Abedus*, the eggs are laid on the backs of males, presumably for protection, and a corixid (*Ramphocorixa*) regularly lays its eggs on a crayfish in pasture ponds. The eggs may be stalked, in Corixidae; spindle-shaped, in Hydrometridae; oval- or barrel-shaped, in Pentatomidae, and may be laid in clusters, by *Zelus*, or singly. The female "broods" or "guards" the eggs in *Nerthra* and some Pentatomidae.

Embryonic development is endoblastic and is not very different from that of other insects. Glandular organs, the pleuropodia, occur on the first abdominal segment and produce the embryonic cuticle. In Polyctenidae the pleuropodia function as a pseudoplacental organ. The incubation period varies from a few days to several months, and development is arrested in species that overwinter in the egg stage as some Miridae. As development proceeds, red eyespots become visible through the chorion. The eggs of many Miridae are embedded in woody tissue and swell because of growth and absorption of moisture, forcing out a yolk plug (*Notostira*).

Eclosion, or hatching, occurs either by splitting or tearing the chorion or by raising the lid. A postembryonic molt takes place during eclosion, and the exuviae are commonly seen attached to the egg shell.

First instar nymphs are active or, in Pentatomidae, sedentary and gregarious. They take little or no food, the yolk remaining from the egg stage being sufficient, in most cases, to carry them through to the second instar. Bacteria are taken up from crypts amidst the egg shells or from the egg surface in Pentatomidae, thus infecting the gastric caeca.

Wing pads begin to show on the hind margins of meso- and metanota at the third stage, the third nymphal instar. Nymphs, also called larvae or neanides, commonly undergo five instars but rare cases have been reported with four, in some Veliidae and Cimicidae. One case, *Dindymus sanguineus* (Fabricius), Pyrrhocoridae, has been reported with nine instars. The final molt and metamorphosis to the adult stage results in an increase in tarsal segmentation, from one to two or from two to three, an increase in number of antennal segments in some cases (4–5 in Pentatomidae), and development of the wings and external genitalia. Also, the ocelli are fully developed at this time, and numerous internal changes occur.

HABITATS OF HEMIPTERA

The Hemiptera occupy various terrestrial and aquatic habitats. Some species are intimately associated with plants and animals.

Table 2. Habitats of Hemiptera

Habitat	Families
Terrestrial	
On surface beneath stones and wood	Nabidae, Reduviidae, Lygaeidae, Enicocephalidae, Leptopodidae
In rotting plant material	Schizopteridae, Lygaeidae
Beneath surface of ground	Cydnidae
Aquatic	
True aquatics	Corixidae, Nepidae, Belostomatidae, Notonectidae, Pleidae, Helotrephidae, Naucoridae
Surface bugs	Gerridae, Veliidae, Hydrometridae, Mesoveliidae
Shore bugs	Gelastocoridae, Ochteridae, Hebridae, Mesoveliidae, Saldidae, Leptopodidae, Dipsocoridae
Marine intertidal	Aepophilidae, Saldidae
Other associations	
In bird nests	Anthocoridae, Cimicidae, Reduviidae, Lygaeidae
Ectoparasites	Polyctenidae
In termite colonies	Termitaphididae, Aradidae, Reduviidae
In spider webs	Microphysidae, Nabidae, Reduviidae
In fungi	Aradidae, Plataspidae
In seeds	Lygaeidae
In stems, foliage, flowers	Anthocoridae, Miridae, Nabidae, Reduviidae, Phymatidae, Tingidae, Lygaeidae, Berytidae, Piesmidae, Pyrrhocoridae, Coreidae, Pentatomidae, Cydnidae, Urostylidae, Phloeidae, Plataspidae

SUBORDERS OF THE HEMIPTERA

Important families of the Hemiptera and their *distinguishing characteristics are treated under the* suborders. Many of the economically important species are given in Table 3.

Hydrocorisae. This suborder contains nine families of water bugs with concealed antennae and without a bulbus ejaculatorius in the male. Many species are predaceous.

Corixoidea. This is a superfamily which contains the single family Corixidae, or water boatmen.

Table 3. Economically Important Hemiptera

Name	Damage
Cimicidae	
Human bedbug (*Cimex lectularius*)	Sucks blood in human dwellings Eggs with small caps, glued in cracks
Tropical bedbug (*Cimex hemipterus*)	Sucks blood in houses, lives in cracks and sleeping mats
Swallow bug (*Oeciacus vicarius*)	In mud nests of swallows, reported on man in houses with swallows
Mexican chicken bug (*Haematosiphon inodora*)	Sucks blood in chicken roosts, eggs laid in cracks
Reduviidae	
Bloodsucking conenose bugs (*Triatoma* spp., *Rhodnius* spp., *Panstrongylus* spp.)	Several species live in nests of wood rats (*Neotoma*), suck blood and transmit Chagas' disease in South America
Miridae	
Tarnished plant bug (*Lygus lineolaris*)	Eggs inserted in plant tissue Overwinters as adult Feeding deforms leaves and stems Punctures cause "cat-facing" on fruits
Lygus bugs (*Lygus hesperus*)	Several species stunt alfalfa, also attack beans, cotton, and other plants They overwinter in adult stage and lay eggs in plant tissue
Apple red bug (*Lygidea mendax*)	Feeding punctures cause pitting of fruits Winters in egg stage, with the eggs inserted in bark
Ash plant bug (*Neoborus amoenus*)	Feeding stunts new growth in spring Eggs laid in stems and bark scars in summer—overwinter in egg stage
Suckfly (*Cyrtopeltis notatus*)	Feeding and egg laying in ring around stems of tomato and tobacco Causes tomato fruits to drop
Cotton flea hopper (*Psallus seriatus*)	Sucks cotton squares causing shedding and whiplike growth of plant Overwinters in egg stage
Garden flea hopper (*Halticus bracteatus*)	Feeding punctures spot leaves of many garden plants Overwinters as adult
Tingidae	
Azalea lace bug (*Stephanitis pyrioides*)	Introduced from Japan. Overwinters in egg stage Three broods a year in New Jersey
Rhododendron lace bug (*Stephanitis rhododendri*)	Overwinters in egg stage Two broods in New Jersey Feeding punctures and fecal spots on undersides of leaves
Basswood lace bug (*Gargaphia tiliae*)	Overwinters in adult stage Eggs laid in spring, partially inserted on undersides of leaves
Eggplant lace bug (*Gargaphia solani*)	Hibernates in adult stage Leaves spotted and yellowed Up to six generations a year in Virginia
Chrysanthemum lace bug (*Corythucha marmorata*)	Overwinters as adult. Eggs inserted deep in veins under leaves Feeding punctures and fecal spots
Sycamore lace bug (*Corythucha ciliata*)	Adults hibernate under bark. Feeding punctures and fecal spots on under surface of leaves
Lantana lace bug (*Teleonemia scrupulosa*)	Feeds on leaves of lantana, causing leaf-drop Introduced into Hawaii from Mexico to control lantana
Ash lace bug (*Leptopypha minor*)	Overwinters in adult stage. Eggs, feeding punctures, fecal spots, and cast skins on under sides of leaves
Lygaeidae	
Chinch bug (*Blissus leucopterus*)	Attacks corn and small grains in the Middle Western states, causing wilting and drying of plants Overwinters in adult stage with 2–3 generations per year
Hairy chinch bug (*Blissus hirtus*)	Attacks lawns in eastern U S, killing grass in spots Life cycle similar to *B leucopterus*
False chinch bug (*Nysius ericae*)	Several species of the genus *Nysius* breed on weeds and migrate in large numbers to cultivated crops, causing wilting
Coreidae	
Boxelder bug (*Leptocoris trivittatus*)	Overwinters as adult Attacks seeds of boxelder, two generations a year
Squash bug (*Anasa tristis*)	Attacks most cucurbits, wilting leaves or killing small plants Overwinters as adult The small brown eggs are laid in groups on the plants in the spring and early summer
Pyrrhocoridae	
Cotton stainers (*Dysdercus suturellus* and *Dysdercus minulus*)	In Florida and Arizona, puncture bolls and seeds and stain the lint
Pentatomidae	
Brown stink bug (*Euschistus servus*)	Breeds on various weeds and attacks most cultivated crops. Feeding stains cotton, pits or causes cat-facing of pears, peaches, and tomatoes, and other plants Overwinters in *the adult stage*
Dusky stink bug (*Euschistus tristigmus*)	
One-spot stink bug (*Euschistus variolarius*)	
Conchuela (*Chlorochroa ligata*)	
Say stink bug (*Chlorochroa sayi*)	
Green stink bug (*Acrosternum hilare*)	
Southern green stink bug (*Nezara viridula*)	
Harlequin bug (*Murgantia histrionica*)	Feeds on foliage, causing wilt and death of the plants. Attacks cabbage and many other garden plants Breeding is more or less continuous, with [illegible] and [illegible] hibern [illegible]

Corixidae lack ocelli, have three dorsal abdominal scent-gland openings as nymphs, and have a unique type of mouthpart. The rostrum is broad, with a pseudosegmentation, and the food consists of algae, as well as small animal forms gathered by the comblike front tarsi, or palae. Respiration is through an air bubble obtained at intervals by touching the surface with the pronotum. Corixids swim with the dorsal side uppermost, using the oarlike middle and hind legs. Eggs are stalked and there are five nymphal instars. Adults sometimes fly to lights in great numbers. In Mexico, the eggs are harvested from lakes to make a bread (ahuatle) and the adults (mosco) are dried and sold in the United States as food for pet birds and turtles.

Nepoidea. The Nepidae, or water scorpions, have a long breathing tube at the tip of the abdomen, through which they obtain air directly from the surface. The front legs are chelate, and the beak is short and stout to suck the juices of other insects on which they prey. Two common types of water scorpion are the long, slender *Ranatra*, of worldwide distribution, and the broad *Nepa*, of the eastern United States and Europe. The eggs are embedded in plant tissue and have two flaplike appendages.

Belostomatidae, or electric-light bugs, are related to the Nepidae but have short, straplike respiratory appendages at the tip of the abdomen. Giant water bugs of the genus *Lethocerus*, reaching 4 in. in length, fly to lights at night. They are pests in fish ponds where they attack fry. They can inflict a painful bite when handled carelessly by man. In China, the adults, called kwai fa shim, are boiled in oil and sold in markets as human food. Eggs are laid on cattails and other plants at the water's edge. Control in fish ponds is by draining or by rearing fry and fingerlings in screened pools. Smaller, 1- to 2-in. belostomatids (*Belostoma*, *Abedus*) lay eggs on the backs of the males, possibly for protection.

Notonectoidea. The Notonectidae or backswimmers, as the name implies, swim ventral side uppermost, the hindlegs serving as oars. Breathing is facilitated by an air bubble, obtained by touching the tip of the abdomen to the surface. *Notonecta* is [illegible] around the spiracles and maintain hydrostatic balance in the water, thus being truly limnetic forms.

Pleoidea. These minute [illegible] dae and He[illegible] fitted for row[illegible] have a single [illegible] gland opening on the abdominal tergum.

Naucoroidea. The creeping water bugs are sometimes separated into the Naucoridae and Aphelocheiridae. They are suboval in body form, with chelate front legs, and with two nymphal scent gland openings widely separated at the hind margin of the third abdominal tergite. Respiration is either by an air bubble or by means of a plastron. In the Old World genus *Aphelocheirus*, the plastron consists of an ultramicroscopic hair pile (2,000,000 hairs per square millimeter), which acts as a physical gill; oxygen diffuses through it from the water to the spiracles of the bug. *Aphelocheirus* is thus able to remain submerged permanently, inhabiting the bottoms of rivers with swift-flowing, well-oxygenated water.

Gelastocoroidea and Ochteroidea. The Gelastocoridae, or toad bugs, and the Ochteridae (or Pelogonidae) are shore-line or mud inhabitants. Both have ocelli. The former are cryptically colored, resembling the sand or mud background. Their eyes are large and their antennae concealed. Ochterids are black with a silky sheen and the antennae, unlike those of other Cryptocerata, are visible from above, though short and partly concealed beneath the eyes. Eggs are laid in sand and mud.

Amphibicorisae. This suborder contains surface water bugs with antennae exposed and without a bulbus ejaculatorius in the male. Spermatheca have a fecundation canal.

Gerroidea. Only the single, diverse superfamily has been proposed for the surface water bugs. All have conspicuous antennae, and the body is clothed with hydrofuge hairs. Three pairs of sensory hairs, the trichobothria, are inserted between the eyes. All Gerroidea are predaceous.

Gerridae are the large water striders with long middle and hind legs and a median scent gland opening, the omphalium, on the metasternum. The claws are inserted before the tips of the tarsi. Most species of *Gerris* are pond or slow-moving stream inhabitants, but *Metrobates* lives on large rivers and *Halobates* inhabits lagoons in tropical seas and even the open ocean. The marine forms are always wingless but all others are polymorphic, with fully winged forms occurring together with short-winged or apterous types. *Halobates* feeds on small anemones and other floating organisms and lays its eggs on feathers, pumice, and presumably anything else that floats. One species, *H. micans* Eschscholtz, occurs in all tropical seas but the remaining 40 or more species are Indo-Pacific.

Veliidae are small water striders which have shorter legs and a longitudinal groove between the eyes. The claws are preapical. Like the Gerridae, these are pond inhabitants (*Microvelia*), stream-riffle bugs (*Rhagovelia*), and marine types (*Halovelia*), but the latter are found only near shore in tropical reefs.

Hydrometridae are long, slender marsh treaders in which the head is longer than the thorax. The claws are apical. These have been reported as predaceous on *Anopheles* mosquito larvae and attempts have been made to transport them for purposes of biological control.

Mesoveliidae and Hebridae are two small families which differ from others in having the well developed ocelli and the single dorsal abdominal scent gland openings of the nymphs. *Mesovelia*

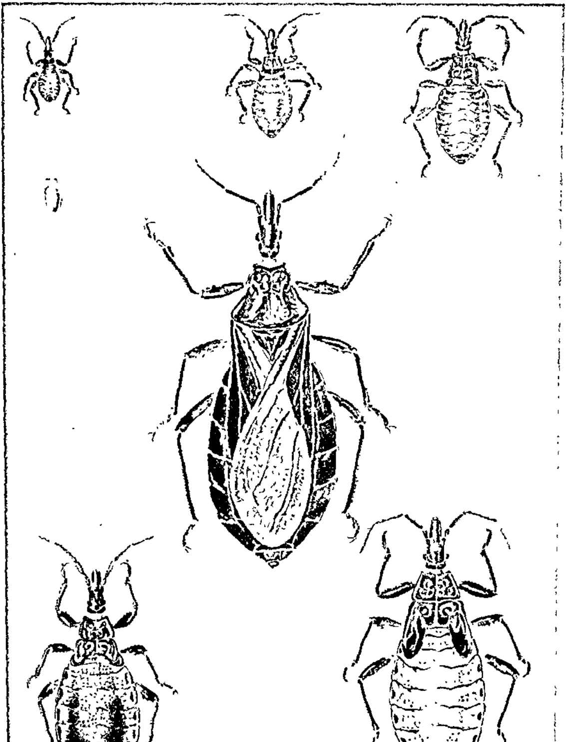


Fig. 6. *Triatoma protracta* (Uhler), egg, five nymphal instars, and adult.

mulsanti White is a swift water strider, but some other species are shore-line inhabitants, as are the Hebridae or velvet water bugs.

Geocorisae. This suborder contains the land bugs with conspicuous antennae and an ejaculatory bulb in the male reproductive system. This suborder includes the majority of Hemiptera. The superfamilies have been divided into two groups, the Cimicomorpha and Pentatomomorpha, with the included superfamilies having certain common characteristics. The Cimicomorpha includes the Cimicoidea, Reduvioidea, and Tingoidea which have distinctly operculate eggs and are without a median spermatheca; the accessory salivary glands are vesicular, the abdomen is without trichobothria, and claws are without discrete lateral arolia. The Pentatomomorpha contains the Pentatomoidea, Coreoidea, Pyrrhocoroidea, and Lygaeoidea. In these superfamilies, the eggs are nonoperculate, a median spermatheca is present, accessory salivary glands are tubular, the abdomen has trichobothria, and the claws have well-developed lateral arolia. Superfamilies that do not fit the above grouping are the Saldoidea, with three pairs of head trichobothria like Amphibicorisae; the Aradoidea, which are like the Pentatomomorpha but without trichobothria; and the Enicocephaloidea and Dipsocoroidea, of doubtful affinities.

Saldoidea. These are the shore bugs which have three pairs of trichobothria on the vertex and a single nymphal scent gland opening. Saldids lay their eggs amidst moss or stones at the edges of streams and ponds. One genus, *Chiloxanthus*, lives on the tundra and overwinters for nine months with hundreds of feet of permafrost beneath and winter ice above. An intertidal relative of the Aepophilidae [illegible]

[illegible] and the Dipsocoridae are predators on small insects under bark or in rotten wood (*Ceratocombus*) or amidst *stones at* the edge of streams (*Cryptostemma*).

Cimicoidea. This is the largest superfamily, with several anomalous *groups* that cannot be placed with certainty at this time. Included in this superfamily are the *predaceous* Joppeicidae of the eastern Mediterranean region, Vianaidae of South America, and Thaumastocoridae *of Australia* and the New World tropics, including the Royal Palm [illegible]

[illegible] and in the nests of birds. The ocelli are distinct, and the front wings *have a* marginal fracture, the cuneus. Fertilization of the female is hemocoelic through special *organs*. The male right genital clasper is large and bladelike; the left is lacking.

Cimicidae contains the bedbugs which probably were derived from the Anthocoridae but differ in habits, in the absence of ocelli, and in having the short wings reduced to pads. Fertilization is usually through an organ of Ribaga on the second, third, or fourth ventrites or the third, fourth, or fifth tergites. The food consists of blood of birds and mammals. The common bedbugs of man are *Cimex lectularius* L., in temperate regions and large cities, and *Cimex hemipterus* (Fabricius), in the tropics. *Oeciacus* lives on swallows in Europe and North America; *Paracimex*, on cave swiftlets in East Asia; *Ornithocoris* and *Haematosiphon*, on chickens in the Western Hemisphere; and *Cimex lectularius columbarius* Jenyns, on pigeons in Europe. Most Old World species, including those of the genus *Cimex*, live on bats, and a few bat parasites are found in the New World. Bedbugs breed continuously in houses and thus are able to complete several generations in a year. Bedbugs and chicken bugs are controlled by residual sprays of 5% DDT or by fumigation. Resistance has been reported to standard dosages of DDT.

The Polyctenidae are bat ectoparasites which resemble the Cimicidae but lack eyes and have ctenidia and strong claws. Reproduction is by hemocoelic fertilization, and the embryo is nourished by a pseudoplacental organ, resulting in true viviparity.

The Miridae family contains a majority of the species of Hemiptera. Included are plant bugs of both herbivorous and predaceous type. Ocelli are lacking, a cuneus is present, the antennae are 4-segmented, and there is only one dorsal abdominal scent gland opening in nymphs. The eggs are embedded in plant tissue. Of the predators, *Cyrtorhinus mundulus* (Breddin) was introduced into the Hawaiian Islands and is credited with controlling the sugar cane leafhopper by sucking its eggs. Plant pests include the tarnished plant bug, *Lygus lineolaris* (Palisot de Beauvois), and other Lygus bugs on alfalfa, cotton, and beans. Control may be obtained with 5% DDT dust applied not later than 30 days before harvest. Lygus bugs overwinter in the adult stage and may undergo two or more generations in a year. Most plant bugs, however, overwinter in the egg stage and are univoltine.

Plant bugs and also some stink bugs and others leave their preferred weed hosts as the growing season advances to suck developing fruit such as apples and pears, thus causing deformity or "catfacing." Control is by clean cultivation or by DDT dust or spray.

Two small families, the Isometopidae and Microphysidae, are close relatives of *Miridae*. Both are predaceous, the former having enormous eyes, and the latter usually having 3-segmented beak and distinct ocelli. Some microphysids live in the webs of spiders and embiids.

Nabidae are the long "damsel" bugs which are slender predators on other insects. The ocelli are well developed, and the rostrum is 4-segmented. The eggs are embedded in plant tissue. Nymphs

have three abdominal scent gland openings. *Nabis ferus* (L.) is a cosmopolitan predator in grasses and other vegetation. The tropical Velocepedidae are related but differ in the broader form and darker coloration.

Enicocephaloidea. This is a unique group in which the head is bilobed, the pronotum trilobed, and the wings completely membranous. There is no prosternal stridulatory groove, the eggs are without an operculum, and the males swarm like gnats. The life cycle is not fully known, but they live under stones, beneath bark, in leafmold and are predators. Some species are wingless; others can cast off their wings.

Reduvioidea. This diverse group includes the assassin bugs or conenose bugs of the family Reduviidae and the ambush bugs of the family Phymatidae. Nearly all have a stridulatory furrow on the prosternum which is scraped by the rostrum to produce a squeaking sound. Ocelli are present, the beak is 3-segmented, and the lateral scent gland openings usually lie beneath the hind coxae rather than on the metapleura. The corsair, *Rasahus*, is large with a spot on the membrane of the wings. It lives beneath stones and feeds on insects but flies to light where it encounters and often bites man. Kissing bugs, the Triatominae, have a longer, more slender rostrum and lack scent glands on the nymphal abdominal tergites. They are exclusively bloodsuckers, living mostly in the nests of wood rats of the genus *Neotoma* in the United States. Several species now regularly live on man in the American tropics: *Panstrongylus megistus* (Burmeister), *Rhodnius prolixus* Stål, and *Triatoma infestans* (Klug). Up to 30% of the bugs are com-

but is transmitted through the feces. Fecal contamination may be at the site of the bite or in the conjunctiva of the eye—the bites are often about the face (hence the name kissing bug) and the feces are rubbed into the eye. The trypanosomes affect muscles of the heart and other organs. Kissing bugs may be controlled with residual sprays, such as DDT, by eliminating nesting rodents or by screening bedrooms. *Rhodnius prolixus* Stål is maintained as a laboratory animal for studies of insect physiology. Other reduviids include the thread-legged Emesinae that live in spider webs and elsewhere, the large wheel bug, *Arilus*, of the southern United States, and many other predators on injurious insects. Phymatids, with long, sometimes concealed antennae, lie in wait in flowers and attack bees and other insects.

Tingoidea. This superfamily contains the lace bugs of the family Tingidae which have wings with many lacelike areolae. They lack ocelli, have four segments in the antennae and beak, and commonly have one or more bulbous or hoodlike elevations on the thorax. Pests are known on ornamental plants such as Christmas berry (*Corythucha*), ash (*Leptoypha*), and other ornamental plants, and some species attack agricultural crops, such as *Gargaphia solani* Heidemann on egg plant. The latter is controlled with Malathion as a 5% dust at 20–30 lb/acre. Lace bug injury is characteristic. The leaves of infested plants have white areas due to feeding and the under surfaces of leaves have black fecal spots and usually cast skins adhering to the leaf hairs. Lace bugs commonly overwinter as adults and pass through more than one generation each year. *Teleonemia scrupulosa* Stål was introduced into the Hawaiian Islands from Mexico in an effort to control the lantana plant.

Aradoidea. Flat bugs of the family Aradidae and their specialized relatives, Termitaphididae, have the mandibular and maxillary setae coiled in the clypeus. They lack ocelli and have 4-segmented antennae. Most, if not all, are fungus feeders. The Termitaphididae have no wings and the margins of the body have numerous small, setigerous lobes. They are known from the nests of termites in the Old and New World tropics. Aradidae are nearly cosmopolitan and may be fully winged, brachypterous, stenopterous, or apterous. Some species are polymorphic, the males having fully developed or stenopterous wings and the females being short-winged.

Lygaeoidea. This is the first of the superfamilies with ventral trichobothria on the abdomen. These erect bristles are in groups of three on either side near the middle of the third and fourth segments and lateral on the fifth, sixth, and seventh segments. The antennae are 4-segmented, as is the beak. Ocelli are present. There are only four simple veins in the membrane. The Lygaeidae include the chinch bug, *Blissus leucopterus* (Say), the false chinch bugs, *Nysius* spp., and the milkweed bugs, *Oncopeltus*. The latter is reared as a laboratory animal for experimental purposes, using milk-

trash. Control is by spraying with Dieldrin (1½ lb/gal of water at 1½ pt/acre) or Toxaphene (2 lb/acre), applying to bases of plants where the bugs congregate. Barrier strips of Dieldrin may also be used in strips 4 rods wide.

Small families related to the lygaeids are the thread-legged Neididae (=Berytidae), the small Piesmidae, formerly included in Tingidae, and the tropical Colobathristidae. Some of the latter have a unique method of stridulation with a filelike arch on the sides of the head.

Pyrrhocoroidea. This group includes the cotton stainers (*Dysdercus*) which attack cotton bolls in the southwestern United States and over most of the tropics, and stout, dark bugs with a reddish border (*Largus*). Pyrrhocorids resemble lygaeids and coreids but lack ocelli. *Dysdercus* may be controlled by a 5% DDT spray.

Coreoidea. The squash bugs and their relatives include the Coreidae, *Anasa tristis* (De Geer); Rhopalidae, *Rhopalus* and *Leptocoris*; Aly...

Leptocorisa; and the Hyocephalidae. They have 4-segmented antennae, a beak, distinct ocelli, and many veins in the membrane. Coreids resemble Lygaeids in arrangement of the trichobothria. The squash bug overwinters as an adult. Eggs are laid on squashes and other cucurbits. There are five nymphal instars and usually only a single drawn-out generation per year. Control is either by DDT 5% dust applied to bugs on soil or under trap boards or by nicotine sulfate in 40% solution at 1½ pt/acre on foliage.

Pentatomoidea. This large group has marginal trichobothria. The antennae are usually 5-segmented and the beak is 4-segmented.

Scutelleridae, or shield bugs, are not injurious in the United States but *Eurygaster* is a pest of grains in Russia, and some metallic green species are plant pests in the tropics.

Cydnidae include the ground burrowing bugs (*Amnestus*), which attack strawberries in sandy soil, and the negro bugs (*Corimelaena*).

Smaller families of little or no economic importance in the United States are the large tropical Tessaratomidae, the Acanthosomatidae, the Dinidoridae, the Aphylidae, the Urostylidae, the bark-like Phloeidae, and the shining, oval Plataspidae.

The true stink bugs, Pentatomidae, include the black and red harlequin cabbage bug, *Murgantia histrionica* (Hahn), and several green stink bugs, *Thyanta*, *Euschistus*, and *Chlorochroa*, on cotton and other crops. Control is by DDT as a dust or spray at 1¼ lb/acre on the foliage or, on cotton, BHC or Dieldrin dust at ½ lb/acre. Stink bugs overwinter as adults; they may produce two generations in a season. Feeding punctures on pears, which are attacked as the cover crop dries, cause cat-facing. One group, the Asopinae, includes predators (*Podisus*) on caterpillars. *See* INSECTA; *see also* ENTOMOLOGY, ECONOMIC. [R.L.U.]

Hemizonida

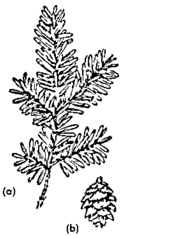
A Paleozoic order of Asteroidea embracing six families. The ambulacral groove is well defined by adambulacral ossicles, but the marginal plates either are not developed or are restricted to the interbrachial angles. The madreporite is either on the upper or the lower surface. *See* ASTEROIDEA; ECHINODERMATA FOSSILS. [H.B.F.]

Hemlock

The genus *Tsuga* of the pine family, characterized by flattened needles with two white lines beneath the needlelike leaves which have distinct short stalks. The cones are small and pendent. Eastern hemlock, *T.* [illegible]

[illegible]

leaves, some of the smaller ones growing upside down, are characteristic of this species. The wood is hard and strong, and is used for construction, boxes, crates, and paper pulp. The bark is one of the principal domestic sources of tannin. It is a common ornamental tree. The annual lumber cut is about 500,000,000 board ft.



Eastern hemlock, *Tsuga canadensis.* (a) Branch. (b) Cone. (USDA)

Carolina hemlock, *T. caroliniana*, a species found in the southern Appalachians, has entire needles and is sometimes grown as an ornamental.

The western hemlock, *T. heterophylla*, which attains a height of 200 ft or more, grows in the extreme Northwest and in Alaska. Its needles resemble those of the eastern hemlock, but the white lines beneath are not so distinct. The annual lumber cut is slightly less than 1,000,000,000 board ft, of which the greater part comes from Washington. An important lumber tree, its uses are similar to those of the eastern species. Another western species, *T. mertensiana*, is of lesser importance commercially. *See* FOREST AND FORESTRY; TREE. [A.H.C.]

Hemoglobin

The oxygen-carrying molecule of vertebrate red blood cells. A molecule of hemoglobin is made of globin, which is a protein, and heme, which is an iron-protoporphyrin complex. Two pairs of polypeptide chains and four hemes are arranged symmetrically about a dyad axis in the molecule, which has the approximate dimensions 55 × 55 × 70 angstroms (A). The iron content of hemoglobin is 0.338%, and its molecular weight is 66,000. *See* PEPTIDE; PROTEIN.

Both heme and globin are synthesized in nucleated red cells of the bone marrow and in reticulocytes, the youngest red cells of the peripheral circulation. The mean life of the hemoglobin molecule in the circulation is that of the red cell, normally about 120 days. When the red cell is destroyed, the porphyrin of heme is degraded to bile pigments. Iron from heme and amino acids derived from the degradation of globin are available for use in metabolic processes which include synthesis of new hemoglobin molecules. *See* PORPHYRIN.

Hemoglobin is the major protein component of blood, normal adult males and females at sea level having about 16 and 14 g/100 ml of blood, respectively. Its normal concentration in red cells is about 34 g/100 ml of packed red cells, and its average amount per normal-adult-human red cell

is about 29×10^{-12} g. Significant deviations from these values occur under altered conditions of red-cell production or destruction; therefore, determinations of these quantities are useful in the differential diagnosis of hematologic disorders.

Probable structure. Although various details of the structure of hemoglobin are known, the complete structure is still to be worked out.

The iron atom in heme is bonded to the four nitrogen atoms of protoporphyrin. It probably forms a fifth bond with an imidazole side chain of histidine, one of the amino acid constituents of the polypeptide chains of globin, to connect heme to globin. The sixth bonding site of the iron atom is available for reactions with external agents. The physiologically active form of hemoglobin contains iron in the ferrous state and is called hemoglobin, deoxygenated hemoglobin, reduced hemoglobin, or ferrohemoglobin. This form combines reversibly at each iron with oxygen to form oxyhemoglobin and with carbon monoxide to form carboxyhemoglobin, which is also known as carbon monoxide hemoglobin. Its affinity for carbon monoxide is much greater than for oxygen; therefore, inhalation of a relatively low concentration of carbon monoxide may result in a high proportion of carboxyhemoglobin and thus may interfere markedly with oxygen transport. Each of the compounds of hemoglobin has a characteristic visible absorption spectrum. The dark bluish color of venous blood and the bright red color of arterial blood are due to the predominance of deoxygenated hemoglobin and oxyhemoglobin in the respective circula-

with oxygen or carbon monoxide but forms stable compounds with ions such as fluoride, cyanide, and azide. High concentrations of methemoglobin in blood, a condition known as methemoglobinemia, are seen in nitrite poisoning, reactions to some drugs, and inherited conditions in which natural reducing mechanisms of the red cell are impaired.

Properties. Several properties of hemoglobin are related directly to its function as a respiratory protein. Reaction of oxygen with the four hemes of each molecule, not as independent entities, but as interacting combining sites, results in an equilibrium curve which is sigmoid in shape. This equilibrium behavior is favorable for unloading a relatively high proportion of hemoglobin-bound oxygen in the tissues. Rapid attainment of equilibrium in the capillaries of the lung and peripheral tissues is achieved by the extreme rapidity with which the reactions of combination and dissociation with oxygen occur. Associated with each heme is an acid group which is stronger in oxygenated than in deoxygenated hemoglobin. Binding of hydrogen ion by this group in the deoxygenation reaction promotes the conversion of tissue carbon dioxide into bicarbonate ion, which can be transported in blood in higher concentration than carbon dioxide. Dissociation of hydrogen ion in the oxygenation reaction results in the formation and release fr

of carbon dioxide in the lungs. Thus, hemoglobin not only takes oxygen into the tissues, but also plays an important part in removing one of the major products of oxidation.

In other species. Although hemoglobin molecules of different species are quite similar in size, reactions, and absorption spectra, they differ in other properties, such as solubility, isoelectric point, and amino acid composition. Differences in molecular composition and properties also occur within a species. Each species has a fetal and an adult form of hemoglobin, and some species have two or more forms of adult hemoglobin which are inherited in accordance with Mendelian laws. Some of the 15 or 20 inherited forms known to occur in man are associated with anemias, the most common of which is sickle-cell anemia. *See* HUMAN GENETICS.

Erythrocruorin is the oxygen-carrying hemoprotein of some worms. This molecule is present in cells in some species and in the body fluid in others, and its molecular weight varies widely among different species. *See* RESPIRATORY PIGMENTS.

Myoglobin. Myoglobin, which occurs in muscle cells, is believed to function in the storage of oxygen. The reactions and absorption spectra of the compounds of myoglobin are similar to those of hemoglobin. A molecule of myoglobin is one-quarter the size of a molecule of hemoglobin; however, myoglobin is not a subunit of hemoglobin since these proteins differ from each other in amino acid composition. [H.I.]

Hemophilia

A rare, hereditary blood disorder marked by a tendency toward bleeding and hemorrhages. Almost entirely restricted to the male members of a family, it was once called the Royal Disease since several European royal lineages were affected. It is transmitted as a sex-linked Mendelian recessive trait and usually passes from an affected male through an unaffected daughter to appear again in a grandson. *See* HUMAN GENETICS.

Actually many varieties exist, but all involve defects of the clotting mechanism, especially thromboplastinogen. Administration of fresh normal blood, which contains the needed factors, allows normal clotting to proceed. *See* BLOOD.

Most of the clinical complaints follow either trauma or infection. Even a slight blow may produce severe hemorrhage into the affected body tissues. Bleeding into the joints, skin, and muscles is common, as is bleeding from the mucous membrane of the oral cavity. Internal bleeding of the genitourinary and gastrointestinal tracts also occurs but for some reason intracerebral hemorrhages are infrequent.

Hemorrhage tends to be massive and accounts for the high death rate, especially in young children. Over 50% of these children die during the first few years of life. Later, survival is probab aided by self-protective measures against and infection and by the appropriate use of fusions, even after minor incidents.

Hemophilic bacteria

Bacteria that in earlier times could be grown well only in culture media containing blood. Since all of these organisms are small (0.2–0.4 by 0.7–1.5 microns), nonsporeforming, gram-negative bacilli, and since most of them are nonmotile and produce diseases of the respiratory tract or the eye, they were placed in a single genus, *Haemophilus* (blood-loving). Further studies have shown that various species require blood for different reasons, so the hemophilic bacteria now are distributed into three genera: *Haemophilus, Bordetella,* and *Moraxella* of the family Brucellaceae. The genus *Haemophilus* retains only species that require either hemin (X factor) or phosphopyridine nucleotide (V factor), or both. The organisms in the genus *Bordetella* use blood components only as detoxicants for substances in the earlier culture medium formulas. Charcoal can be substituted for blood, and there are casein hydrolysate media that contain neither but will support the growth of *Bordetella*. The genus *Moraxella* comprises species that do not require either X or V factors but need other growth factors from serum. *See* BACILLACEAE.

H. influenzae, or Pfeiffer's bacillus, is a common inhabitant of the human respiratory tract, where it seldom produces symptoms. It is not the cause of human influenza (*see* INFLUENZA). However, during respiratory infection with other organisms or viruses it may become a secondary invader. This species may show bipolar staining and capsulation. The encapsulated strains are divided into six serologic types (A–F) on the basis of specific capsular polysaccharides. Type B is more pathogenic and less frequent than the others and is an important cause of meningitis in children. *See* MENINGITIS.

H. suis is an encapsulated organism similar to *H. influenzae* but which inhabits the respiratory tract of swine. This bacterium, in association with swine influenza virus, causes swine influenza. There are three serological types. *See* BACTERIOLOGY, MEDICAL; SEROLOGY. [W.F.V.]

Bibliography: R. S. Breed et al. (eds.), *Bergey's Manual of Determinative Bacteriology*, 7th ed., 1957; G. S. Wilson and A. A. Miles (eds.), *Topley and Wilson's Principles of Bacteriology and Immunity*, 4th ed., 2 vols., 1955.

Hemorrhage

The escape of blood from within the vascular system. Hemorrhage may result from either trauma or disease of the vessel wall. In some cases there is no apparent break in the vessel wall, and the blood cells escape by a process called diapedesis.

Causes. The escape of blood following rupture of a vessel wall as a result of trauma is obvious and needs no further explanation. The causes other than trauma can be divided into three main groups.

The first group consists of those conditions in which the[...]

the

for

t = epithelium of renal tubule g = glomerulus

Fig. 1. Hemorrhagic infarct of the kidney, showing epithelium of a renal tubule (t) and a glomerulus (g).

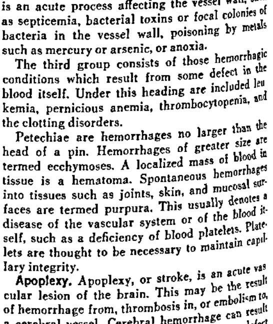
with an elevated blood pressure, can result in a break in the wall and subsequent hemorrhage. An infarct, or tissue death from any cause, may also result in hemorrhage.

The second group is those causes in which there is an acute process affecting the vessel wall, such as septicemia, bacterial toxins or focal colonies of bacteria in the vessel wall, poisoning by metals such as mercury or arsenic, or anoxia.

The third group consists of those hemorrhagic conditions which result from some defect in the blood itself. Under this heading are included leukemia, pernicious anemia, thrombocytopenia, and the clotting disorders.

Petechiae are hemorrhages no larger than the head of a pin. Hemorrhages of greater size are termed ecchymoses. A localized mass of blood in tissue is a hematoma. Spontaneous hemorrhages into tissues such as joints, skin, and mucosal surfaces are termed purpura. This usually denotes a disease of the vascular system or of the blood itself, such as a deficiency of blood platelets. Platelets are thought to be necessary to maintain capillary integrity.

Apoplexy. Apoplexy, or stroke, is an acute vascular lesion of the brain. This may be the result of hemorrhage from, thrombosis in, or embolism to, a cerebral vessel. Cerebral hemorrhage can result from a rupture of the vessel wall, as from a defect



a = alveolar spaces

Fig. 2. Section of lung tissue with hemorrhage within alveoli (a).

in the wall such as a congenital aneurysm, or it may be secondary to death of cerebral tissue as occurs with an infarct. Intracerebral hemorrhages may be small petechial or perivascular hemorrhages, or they may be massive. The small hemorrhages can result from minimal trauma, poisoning such as from carbon monoxide, or from blood diseases such as leukemia. Massive cerebral hemorrhage is frequently the result of trauma or vascular disease. With cerebral hemorrhage or death of cerebral tissue there is a loss of function of the regions involved. One such complication is a paralysis of the muscles of the extremities on the side opposite the lesion. Since the brain is enclosed in a nondistensible box, the skull, the increase in size due to the addition of the blood raises the intracranial pressure.

(b) = lateral ventricle distended with blood

Fig. 3. Intracerebral hemorrhage secondary to the rupture of an intracerebral vessel, showing lateral ventricle (b) distended with blood.

Compensatory mechanisms. Following the sudden loss of a large quantity of blood, the body reacts to compensate for this loss and the blood clotting reaction is accelerated. The media of the vessel wall is retracted and contracted. The endothelium becomes curled and crenated. As the blood volume is decreased the carotid sinus reflex comes into play. This results in a generalized vasoconstriction which affects the arteries, precapillary, and capillary vessels. With a lowering of capillary pressure there is a shift of fluid from the tissue spaces into the plasma. This tends to restore the blood volume but lowers the hematocrit (the relative percentage of erythrocytes, or red cells, in the blood). Following a loss of 500 ml of blood a healthy person can restore his blood volume in 24 hours. The protein content can be restored in a few hours but the replacement of the cellular elements takes days or even weeks. Failure of these mechanisms to compensate adequately can result in secondary shock with dire consequences.

Morphologic changes of hemorrhage. Small hemorrhages can be completely absorbed leaving no tissue alterations; larger volumes of blood act as a foreign substance which is destroyed and removed by the phagocytes. The red cells are broken down, liberating hemoglobin. This released hemoglobin is divided into two moieties, the iron-free hematoidin and the iron-containing hemosiderin.

In animals dying of severe or fatal hemorrhage a state of *severe ischemia of all of the organs is* found. The mucous and serous surfaces are pale; the organs are drier than normal and reduced in weight. [R.A.V.]

Bibliography: W. A. D. Anderson (ed.), *Pathology*, 3d ed., 1957; W. Boyd, *A Text-book of Pathology*, 6th ed., 1953.

Hemorrhagic septicemia

An infectious disease of fowl (fowl cholera), rabbits (snuffles), swine, buffalo, and other animals. It is also known as pasteurellosis and is caused by *Pasteurella multocida*, a small coccobacillus. The disease may be transmitted to man. The microorganism is nonmotile and does not form spores. In films of blood, exudate, or organs the organism is gram-negative and shows bipolar staining, that is, the ends of the organism stain more intensely than the center.

Blood agar medium with cystine will support good growth of the organism at 37°C. The microorganism will grow in a medium containing carbohydrates and will produce acid, but not gas.

Pasteurella multocida, first isolated by Louis Pasteur in 1880, is a commensal, host-specific parasite of either the upper respiratory or the intestinal tract. However, when the host-parasite balance is disturbed, the commensal strains may become virulent and cause acute hemorrhagic septicemia in fowl and domestic animals. Virulent, nonhemolytic, mucoid colonial forms dissociate to form avirulent, rough, colonial forms. In man, infection by this organism has resulted in generalized infection, wound infection after animal bites, and ocular or cerebral manifestations.

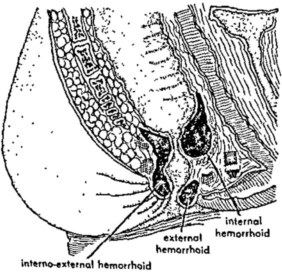
The microorganisms are sensitive in vitro to most antimicrobial drugs. Proper animal husbandry is more effective and economical than treatment or immunization. For taxonomy *see* BRUCELLACEAE; *see also* BACTERIAL MOTILITY; VIRULENCE. [K.F.ME.]

Hemorrhoids

Soft, irregular venous abnormalities occurring in and above the anus. The exact causes are unknown but pregnancy, constipation, physical exertion, and congenital venous abnormality have been considered as etiologic factors.

The thin-walled vessels become engorged and dilated, and thrombosis is frequent. An associated inflammation occurs and there is replacement by scar tissue.

External hemorrhoids lie outside the anal ring and are commonly seen in two forms. The first is the acutely thrombosed hemorrhoid which gives great pain or tenderness that may be dramatically relieved by minor surgery. The second form occurs as part of a more extensive network of varicosities about the anal region. Symptoms vary greatl



Types of hemorrhoids. (After Pennington from W. A. N. Dorland, ed., *American Illustrated Medical Dictionary*, 19th ed., Saunders, 1942)

often there is merely a slight bleeding tendency while at stool

Internal hemorrhoids may be simple, hard-to-detect varicosities lying well within the anal canal, or they may be huge, protruding, blood-filled vessels that give intense pain. Severe hemorrhage may occur as well as the passage of lesser amounts of bright red bl...

...eat- ... re- currence or complications. *See* ANUS; VEIN.

[E.G.ST.]

Hemp

The fiber and the plant, *Cannabis* ... nc ... la ...

ra ... material used in the manufacture of rope, now is used mostly in the production of small yarns, twines, and canvases and, to some extent, in making paper of a special type (*see* FIBER, NATURAL).

The plant is about 6–8 ft tall at maturity and has a stem roughly the diameter of a lead pencil (Figs. 1 and 2).

Fig. 1. A hemp harvester cutting and spreading hemp for retting. (USDA)

Hemp is an annual crop, most of which is produced in Italy, Yugoslavia, southern France, Turkey, and Russia (*see* ANNUAL PLANTS). In the United States, production has gradually declined (except during war years) from more than 10,000,000 lb in 1895 to none in 1958. Hemp seed production in the United States was centered in Kentucky and fiber production in the upper Middle West.

Hemp is planted about the same time as corn and requires about 4 months of frost-free weather to produce a crop of fiber. It needs a fertile soil and good drainage.

The fiber is separated by crushing the retted (partially rotted) stems and beating out the nonfibrous material.

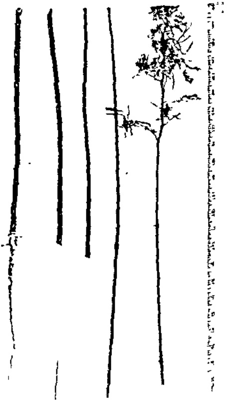


Fig. 2. One hemp plant that has been cut into sections. (Agr. Research Serv., USDA)

Hemp contains the drug marijuana and a permit from the U.S. Treasury Department is required to grow or possess any part of this plant in the United States. *See* MARIJUANA.

[E.C.N.]

Henry

A unit of self- and mutual inductance. One henry is the self-inductance of a circuit or the mutual inductance between two circuits if there is an induced electromotive force of 1 volt when the current is changing at the rate of 1 amp/sec. The henry may also be defined in terms of the number of flux linkages per unit current. The henry is 1 weber-turn/amp. *See* INDUCTANCE; *see also* ELECTRICAL UNITS.

[K.S.M.]

Hepaticae

A class of lower green plants commonly called liverworts that belong to the phylum Bryophyta. Although there is a great diversity of external form, most of the gametophytes (gamete-producing plants) are dorsoventrally differentiated. These plants are considered the most primitive of the existing land plants. Liverworts are widely distributed over the world, but have their greatest diversity in the tropics of the Americas and East Indies.

Except when the plants occur in masses, they are quite inconspicuous and are usually confused with mosses which they resemble somewhat in their external appearance. In the presence of adequate moisture, they grow on soil, rocks, and tree trunks. Usually the plant body is a thin, prostrate thallus sometimes having a short central axis with leaflike appendages. On the lower surface there are rhizoids (rootlike structures) which function in anchorage and absorption.

Reproduction. The sex organs, antheridium (male) and archegonium (female), may be produced on the same plant or on different plants. The sperms, produced in the antheridia, are flagellate and motile. The sporophyte (spore-producing generation), which develops as a result of fertilization, remains on the female gametophyte, and for a time it is parasitic. The spores are developed within a sporangium, or capsule, which lacks a sterile region or columella internal to the spore-producing tissue as occurs in the mosses (*see* MUSCI). After discharge from the sporangium the spores germinate, forming short protonema from which arise new liverwort plants (gametophytes). Fossil remains of liverworts have been found in the upper Carboniferous of England (*see* GEOLOGY). Since the fossils found do not differ significantly from modern liverworts, they are of little value in ascertaining phylogenetic relationships.

[illegible] on [illegible] re [illegible] ually divided into two groups: the Jungermanniales and the Marchantiales.

The Jungermanniales are called the leafy liverworts because of their chlorophyll-containing ribbonlike or leaflike bodies (Fig. 1). The thallus is never differentiated, and it lacks the air spaces and external pores characteristic of the Marchantiales. Only unicellular, smooth-walled rhizoids are present. The sporophyte often produces a long stalk (sporangiophore) bearing the sporangium, or capsule, on the upper end. When mature the sporangium dehisces (opens) by four valves.

The Marchantiales are called thallus liverworts, the best known example of which is *Marchantia polymorpha*. The flat plant body is composed of several distinct layers of tissue, having the chlorophyll-containing cells of the uppermost layer separated by numerous air chambers which are exposed to the external atmosphere through pores. Rhizoids are of two kinds, smooth-walled and tuberculate-walled. The male and female sex organs, each borne on separate plants, are

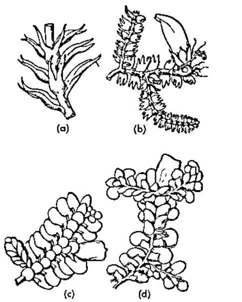


Fig. 1. Jungermanniales. (a) *Herberta*, showing three ranks of equal bifid leaves (*after Muller*). (b) *Lepidozia*, ventral aspect, showing reduced ventral leaves; (c) *Leucolejeunea*, ventral aspect, showing reduced ventral leaves and complicate dorsal leaves; (d) *Radula*, ventral aspect, showing absence of ventral leaves but complicate dorsal leaves (*after A. Lorenz*). (*From E. W. Sinnot and K. S. Wilson, Botany: Principles and Problems, 5th ed., McGraw-Hill, 1955*)

grouped in the tops of special stalked structures known respectively as antheridiophores and archegoniophores.

Alternation of generations. The life cycle of liverworts is well illustrated by *Marchantia* (Fig. 2). The gametophyte, or gamete-producing generation, is the dominant phase. In *Marchantia*, the gametophyte is a flat, dichotomously branched

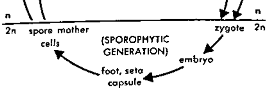


Fig. 2. Diagram of the life cycle of the liverwort *Marchantia*, indicating the stages where chromosome changes occur. (*From H. J. Fuller, The Plant World, rev. ed., Holt, 1951*)

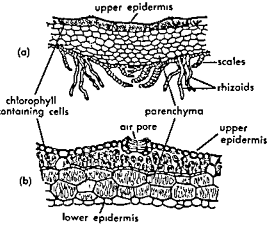


Fig. 3. Transverse sections through the thallus of *Marchantia*. (a) Middle portion with scales and rhizoids on the under side. (b) Margin of the thallus more highly magnified, showing colorless reticulately thickened parenchyma, epidermis of the upper side, cells containing chlorophyll, air pore, and lower epidermis (after Goebel). (*From H. Kraemer, Applied and Economic Botany, 2d ed., Wiley, 1916*)

(forked) thalloid structure produced by apical growth of the thallus. As development proceeds, the thallus produces numerous rhizoids from the ventral (under) surface. The internal anatomy shows considerable differentiation of tissues (Fig. 3). Starting with the upper surface, there is a single epidermal layer, subtended by an area of air chambers which are connected to the external atmosphere through regularly spaced pores or openings (*see* EPIDERMIS, PLANT). The pores are surrounded by four rows of four cells each, the upper and lower circles being somewhat smaller, giving the whole structure a barrel-shaped appearance Below these openings, the vertical pillars of cells and the multibranched filaments at the base of the chambers have numerous oval chloroplasts which give the rich, dark green color so characteristic of the plant. The remainder of the thallus is made up of densely arranged parenchymatous cells (*see* PARENCHYMA). The lower parenchyma tissue has few chloroplasts, but it contains numerous large mucilage cells which probably function as storage tissue.

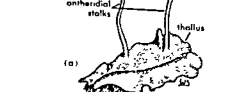


Fig. 4 (a) M...

The plants are dioecious, that is, they have male and female sex organs on different plants. The antheridia, or male sex organs, are produced on stalked, disk-headed branches called antheridiophores which grow upward from the apical notch of the thallus or gametophyte (Fig. 4). The antheridia are embedded in cavities of the disk,

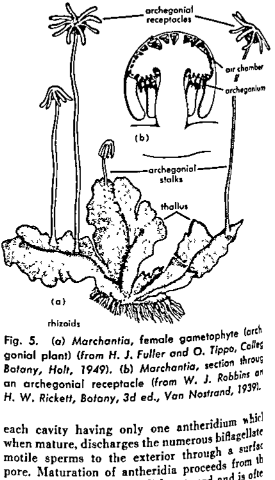


Fig. 5. (a) *Marchantia*, female gametophyte (archegonial plant) (*from H. J. Fuller and O. Tippo, College Botany, Holt, 1949*). (b) *Marchantia*, section through an archegonial receptacle (*from W. J. Robbins and H. W. Rickett, Botany, 3d ed., Van Nostrand, 1939*).

each cavity having only one antheridium which when mature, discharges the numerous biflagellate motile sperms to the exterior through a surface pore. Maturation of antheridia proceeds from the center of the antheridial disk outward and is often indicated by the cell walls becoming deep purple in color.

The archegonia are found on similiar stalks called archegoniophores (Fig. 5). At the apex of each archegoniophore about nine finger-shaped structures project outward like the ribs of an umbrella. The archegonia are formed in groups near the bases of these arching projections. Each projection may produce 12–16 archegonia in succession from its tip inward.

There has long been some question as to how the sperms reach the egg cells in the archegonia. It has been observed that raindrops may splash the sperm as much as 2 ft. Mites have been ob

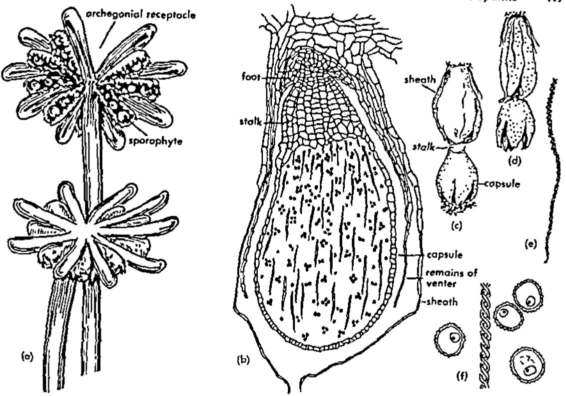


Fig. 6. (a) Mature archegonial receptacles of *Marchantia* bearing sporophytes (*from W. J. Robbins and H. W. Rickett, Botany, 3d ed., Van Nostrand, 1939*). (b) *Marchantia polymorpha* L., longitudinal section of a nearly mature sporophyte; (c,d) surface views of sporophytes during and after shedding of spores and elaters; (e) portion of an elater; (f) spores and portion of an elater (*from G. M. Smith, Cryptogamic Botany, vol. 2, 2d ed., McGraw-Hill, 1955*).

served on the archegoniophore with live sperm on their bodies. Following a rain, or with heavy dew, there is a water film which is probably adequate for the swimming of the sperms from the antheridia to the archegonia.

As a result of fertilization a zygote is formed which develops into a sporophyte (Fig. 6). This is accomplished by a series of divisions of the zygote (fertilized egg) which finally results in a structure having a foot embedded in the gametophyte; an elongated seta, or stalk, which shows its most rapid development at the time of spore maturation; and

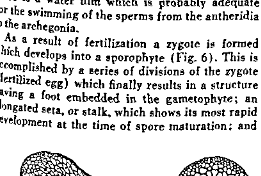


Fig. 7. (a) Portion of surface of *Marchantia* thallus with numerous rhomboidal areas, each area with a central pore. Note gemma cup with gemmae (*from H. J. Fuller and O. Tippo, College Botany, Holt, 1949*). (b) Gemma of *Marchantia polymorpha* L. (*from G. M. Smith, Cryptogamic Botany, vol. 2, 2d ed., McGraw-Hill, 1955*).

a sporangium, or capsule, within which the spores are developed following a meiotic division of the spore mother-cells (*see* MEIOSIS). Special pointed elongated cells, known as elaters, aid in the dispersal of spores by being sensitive to internal water tensions which cause them to coil and uncoil. If the spores fall upon a suitable substrate, they germinate immediately and produce new gametophytes.

In *Marchantia* there are also two methods of vegetative reproduction. The first results from the growth habit of the plant. As it grows and branches, the older portions of the plant decay, thus separating the branches into individual plants. The second method is in the formation of gemma cups on the dorsal surface (Fig. 7). Within each cup several small, notched, lens-shaped bodies are produced which, when detached, grow from both ends simultaneously and develop into new plants. This provides for rapid dispersal during the growing season. *See* TRACHEOPHYTA. [P.A.V.]

Bibliography: *See* BRYOPHYTA.

Hepatitis

... outward from a central vein. Between the ...

are the tiny biliary ductules. Surrounding each lobule is supportive tissue containing larger blood vessels and biliary ducts. Recognition of liver disease depends upon the clinical and pathologic evidence of damage to the various parts of the biliary system, the vascular system, the connective framework, and the liver cells, or parenchyma.

Among the more common types of hepatitis are those of viral origin. Two major forms exist, infectious hepatitis and homologous serum jaundice. The former is caused by a filtrable virus which may cause epidemics, some of which have been traced to fecal contamination of drinking water. Serum jaundice, also caused by a filtrable virus, exists in the blood plasma of many persons and may be passed to another individual through blood transfusion or related means. Although the pathologic lesions and clinical courses are quite similar, one distinguishing feature is the difference in the length of the incubation period, 2–4 weeks for infectious hepatitis and 4–34 weeks for serum jaundice. Early fever, nausea, fatigue, malaise, and other nonspecific complaints may precede a variable degree of jaundice. The clinical course varies considerably in length, severity, and symptomatology. Most cases display a self-limiting, benign nature, but occasionally severe illness or even death quickly follows a rapidly progressive disease.

Other forms of hepatitis include those following some degree of parenchymal damage, ranging from the mild involvement possible in infectious mononucleosis to the severely destructive lesions produced by acute chloroform poisoning and by other drugs or chemicals. *See* INFECTIOUS MONONUCLEOSIS; POISON.

Interstitial hepatitis is an inflammation of the supporting tissue of the liver, often with little or no involvement of parenchyma. Biliary tract obstruction and blood-borne agents are the two large categories of factors which may produce this type of hepatitis.

Other specific agents which may produce some type or degree of liver inflammation are quite numerous, but such examples as syphilis, typhoid fever, malaria, toxemias of pregnancy, nutritional deficiencies, and parasitic infestations indicate the wide range of possibilities. *See* LIVER; MALARIA; NUTRITION; PARASITOLOGY, MEDICAL; PREGNANCY, DISORDERS OF; SYPHILIS; TYPHOID FEVER.

[E.C.ST.]

Heptode

A 7-electrode vacuum tube used as a mixer in a radio receiver. The five grids between the cathode and plate are operated at alternately low and high potentials with grids 1, 3, and 5 at zero or slightly negative potential and grids 2 and 4 at a high positive potential approximately equal to the plate potential. The function of the electrodes in the heptode is as follows: the cathode serves as a source of electron current; the first grid serves to inject signal voltage and as a source of bias for automatic volume control; the second grid is a screen to reduce coupling between the signal and oscillator grids; the third grid serves to inject the local-oscillator voltage; the fourth grid acts as a screen to reduce coupling between oscillator and output circuits; the fifth grid is a suppressor grid to improve plate-current characteristics and further reduce coupling between oscillator and output circuits; the plate acts as the collector of the modulated electron current.

Heptode characteristics are in general superior to hexode characteristics. In a heptode mixer it is possible to use the third grid instead of the first grid for local oscillator voltage injection because of the extra shielding between the third grid and plate introduced by the presence of the fifth, or suppressor, grid. This arrangement allows the first grid to be used for signal injection and makes it possible to obtain a variable-mu action from this grid, which in turn makes good automatic volume control possible. In general, it is difficult to design a tube with a variable-mu characteristic on any but the first grid. *See* VACUUM TUBE. [K.R.S]

Herbarium

A collection of plant specimens, pressed and mounted on paper or placed in liquid preservatives, and systematically arranged for subsequent consultation. Most of the great botanical gardens have such collections (*see* BOTANICAL GARDENS).

Some of the great herbaria of the world (Fig. 1*a*), with approximate number of specimens, are listed below:

Royal Botanical Gardens, Kew	5,000,000
V. L. Komarov Botanical Institute, Leningrad	5,000,000
British Museum (Natural History), London	4,000,000
Berlin-Dahlem, Botanischer Garten (damaged in World War II)	4,000,000
Jardin des Plantes, Paris	3,500,000
Harvard University, Cambridge	3,200,000
Conservatoire et Jardin Botaniques, Geneva	3,000,000
Indian Botanic Garden, Calcutta	2,500,000
U.S. National Herbarium, Washington	2,250,000
N.Y. Botanical Garden, New York	2,211,000
Royal Botanic Garden, Edinburgh	1,500,000
National Herbarium, Melbourne	1,500,000
Missouri Botanical Garden, St. Louis	1,500,000

The chief value of a herbarium lies in the bringing together of large numbers of specimens of plants for comparative studies. It is important for botanists to have a knowledge of the appearance of living plants in their customary habitats, but without collections of preserved plants, such works as monographic treatments of various groups or manuals of the flora of various regions could not be prepared. In the preparation of monographs, researchers often borrow specimens of the respective groups each is studying from numerous herbaria and affix annotation-labels giving the results of their study. Thereafter, these specimens provide standards for the identification of other specimens and consequently have greater value in the herbar-

ium. Students preparing floristic treatments of various physiographic (or political) divisions also find the herbarium indispensable. Often the examination of many thousands of specimens is necessary before accurate statements can be made about the distribution of the plants of a certain group (or groups) within a given region.

Type specimens. These are the most valuable specimens in any herbarium. They are specimens upon which the name of some taxon, such as a species or variety, has been based. Because of their value, type specimens are usually accorded special care by curators, sometimes being kept in separate cases to avoid unnecessary handling.

Specimens for the herbarium need to be collected carefully, and at the time the collection is made, the collector should record the date, the place where collected, and any other pertinent data. The specimens are then placed in a plant press for pressing and drying (Fig. 2). Blotting papers, newspapers, and ventilators are used in the drying process, often expedited by artificial heat.

Preservation of specimens. Mounting of the dried plants on sheets of mounting paper readies them for permanent filing. In the United States these sheets of heavy, good quality paper are of standard size, 11½ by 16½ in. (Fig. 1b). Methods of attaching specimens vary, but perhaps it is best done by coating the under surface of the leaves with glue and then placing the plant on the mounting paper. Heavier stems may be held in place by fastening with gummed strips.

The herbarium label, pasted on the lower right corner of the mounting sheet, is an important addition on which are recorded the scientific name of the plant, the place where collected, the date of collection, and the name of the collector, as the minimum information. The habitat, including altitude above sea level, may also be an important item of this record.

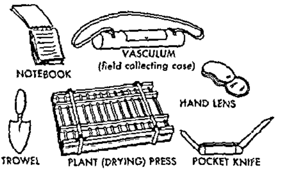


Fig 2. Plant collector's standard equipment.

The herbarium is arranged by families according to one of the commonly used systems of classification. These families may be numbered consecutively, and cards prominently displayed for ready reference show the arrangement. Specimens of each genus are filed in a folded manila genus cover, 16⅝ by 12 in. in size (when folded), with the family name and number, and the name of the genus printed at the lower left corner. The genera may be filed under the families in alphabetical order (the most convenient) or according to some phylogenetic sequence. Sheets representing the various species may be filed in separate genus covers, or those of each species may be placed together in a species cover and then filed in the genus cover. Compared with the genus cover, the species cover is lighter in weight and of slightly smaller dimensions. Species may be arranged within the genus in alphabetical order, or according to some given systematic sequence appearing in a monograph of the group.

The specimen sheets are filed in herbarium cases. The modern herbarium case is of welded and reinforced steel construction, with pigeonholes each usually 19 in. deep, 13 in. wide, and about 8 in. high.

Preservation of specimens from insect depredations is essential. Notable among insect pests are [illegible]

attended for a few months may be largely destroyed. Constant vigilance and the combined use of insecticides and repellents are imperative. Insecticides include hydrocyanic acid, carbon disulfide gas, and dichlorodiphenyltrichloroethane (DDT). Another insecticide is a mixture of three parts ethylene dichloride with one part carbon tetrachloride. Repellents are substances that repel but do not necessarily kill the insects. The two principal repellents are naphtha flakes and paradichlorobenzene (PDB), often placed in small muslin bags within each herbarium case. See PL[illegible] CLASSIFICATION; PLANT KINGDOM. [illegible]

Bibliography: See PLANT TAXONOMY.

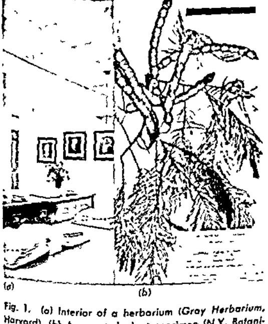


Fig. 1. (a) Interior of a herbarium (Gray Herbarium, Harvard). (b) A mounted plant specimen (N.Y. Botanical Garden).

Herbicide

One of a number of chemicals used to kill weeds, that is, unwanted plants. Use of herbicides has developed chiefly in the twentieth century, slowly at first and then rapidly and dramatically after 1945 when 2,4-D was introduced. A few milestones are: 1896–1900, introduction of selective sprays for controlling broad-leaved weeds in cereal crops; 1919, discovery of the translocation of arsenic in wild morning glory by George Gray in California; 1925, introduction of sodium chlorate for killing weeds by application to the soil; 1934, importation from France to America of sodium dinitrocresylate as the first organic selective weed killer; 1945, advent of 2,4-D, the growth-regulator type of herbicide. Since the introduction of 2,4-D, new chemicals and new methods have come along at an accelerated rate until now the manufacture and use of herbicides has become a multimillion-dollar business annually.

As is true of medical science and of other aspects of pest control, the control of weeds by means of herbicides has benefited practically everyone. The values of agricultural crops have been increased by billions of dollars; forests, rangelands, parks, and recreational areas have been freed of noxious plants; roadsides and industrial areas are kept clean and sightly; and millions of people have been relieved of the suffering caused by pollens and poisonous plants. Looking to the future, these benefits will be multiplied greatly as new and better herbicides are discovered. In fact, the reclamation of hundreds of millions of acres now covered by brush and undesirable trees, not only in the Americas, but in Africa, Asia, and Australia, constitutes the greatest single source of land available for agricultural use.

Classification of herbicides. There are available dozens of chemicals, hundreds of formulations, and a multitude of methods for killing weeds. The following classification of herbicides by method of application has proved extremely useful in discussing weed control problems:

1. Selective foliage contact sprays
2. Selective foliage translocated sprays
3. Selective root applications
4. Nonselective foliage contact sprays
5. Nonselective foliage translocated sprays
6. Nonselective root applications

A great majority of the available herbicides and methods for their use fall naturally into these six groups. The few exceptions will be discussed separately.

Selective herbicides are those that kill some members of a plant population with little or no injury to others. Examples are sulfuric acid, which, at proper strength, will kill wild mustard in barley, or knotweed in young onions.

Nonselective herbicides are those that kill all vegetation to which they are applied. Examples are sodium arsenite solution or aromatic oil. These are finding wide usage as their virtues become more widely recognized. They are used to keep roadsides, ditch banks, and rights-of-way open and weed free. Fire prevention on highways and industrial sites and elimination of hazards at intersections are additional uses. Often these general weed control materials are used to get rid of plants that serve as alternate hosts for insects and plant diseases.

Application methods fall into two natural groups as indicated in the above scheme of classification; these are application to foliage and application to soil for uptake by roots. In regions of frequent summer rains, or by using sprinkler irrigation, both these methods may be used to advantage. By spraying the foliage, leaf absorption may take place. If this is followed by rainfall or sprinkler irrigation the chemical not absorbed by foliage may be washed into the soil and taken up by roots. Sodium chlorate may be used in this way.

Herbicide action. There are many factors that influence the action of herbicides on plants.

Differential wetting. Most leaf surfaces are waxy; many are corrugated or ridged; some are covered by small particles of wax (bloom). Such leaves differ in their retention of spray droplets. Many selective sprays such as copper salt solutions, sulfuric acid, and sodium dinitrocresylate kill weeds selectively because they wet the weeds, but bounce off of the crop plants.

Orientation of leaves. Most grass leaves stand in a relatively vertical position. Many weed leaves are arranged horizontally so that they retain more spray than do grass leaves.

Location of growing points. Growing points and buds of most cereal plants are located in a crown, at or below the soil surface. Furthermore they are wrapped within the mature bases of the older leaves. Hence they may be protected from herbicides applied as sprays. Buds of many broad-leaved weeds are located at the tips of shoots and in axils of leaves. These are much more exposed to spray solutions.

Growth habits. Some perennial crops, such as alfalfa, vines, and trees, have a dormant period in winter. At that time, a general contact weed killer may be safely used to get rid of weeds that later would compete with the crop for water and plant nutrients.

Application methods. By arranging spray nozzles close to the soil level it is possible to kill low growing weeds, but not an upstanding crop. In this way, general contact sprays may be used in corn, sugar cane, or sorghum. In California, such directed spraying is used to kill young grass in cotton with oil, and to kill wild morning glory in beans with 2,4-D.

Protoplasmic selectivity. Just as some people are immune to the effects of certain diseases while others succumb, so some weed species resist the toxic effects of herbicides whereas others are injured or killed. This results from inherent properties of the protoplasm of the respective species. One selectivity of this sort distinguishes between certain weeds and crops on the basis of their enzyme systems. If weeds, having in their cells a β-oxidizing enzyme,

Some important herbicides

Common name or designation	Chemical name	(a) Soluble in: (b) Available as:	Use
CDAA (Randox)	2-Chloro-*N*,*N*-diallyl acetamide	(a) Organic solvents (b) Emulsifiable concentrate	Selective grass killer
PMA	Phenyl mercuric acetate	(a) Water (b) Liquid concentrate	Selective spray against crab grass in lawns
NaTCA	Sodium trichloroacetate	(a) Water (b) Dry powder	Grass killer through soil application
Acrolein	Acrolein	(a) Water (b) Liquid	Aquatic weed killer
Arsenic (white arsenic)	Arsenic trioxide	(a) Insoluble (b) Dry powder	General soil sterilant
Arsenite, sodium	Sodium arsenite	(a) Water (b) Liquid concentrate	General contact spray
MAA (Sodar)	Disodium methylarsonate	(a) Water (b) Dry powder	Selective spray for crab-grass
2,3,6-TBA	2,3,6-Trichlorobenzoic acid	(a) Salts in water, esters in oil (b) Liquid concentrate	Selective translocated spray, soil applied perennial weed killer
Amoben	2,5-Dichloro-3-aminobenzoic acid	Experimental	Selective preemergence herbicide in vegetables and soybeans
Dinoben	2,5-Dichloro-3-nitrobenzoic acid	Experimental	Selective preemergence herbicide in vegetable crops
Biuret	Biuret	(a) Water (b) Dry crystals	Pre- and postemergence herbicide especially toxic to grasses
Borate (penta)	Sodium pentaborate	(a) Water (b) Dry powder	General soil sterilant
Borate (meta)	Sodium metaborate	(a) Water (b) Dry powder	Contact spray and soil sterilant
CDEC (Vegadex)	2-Chloroallyl diethyldithiocarbamate	(a) Organic solvents (b) Emulsifiable concentrate	Selective grass killer
Barbane	4-Chloro-2-butynyl *N*-(3-chlorophenyl) carbamate	(a) Organic solvents (b) Emulsifiable concentrate	Selective postemergence spray against wild oats
Vapam (SMDC)	Sodium *N*-methyl dithiocarbamate dihydrate	(a) Water (b) Liquid solution	Soil fumigant
CIPC	Isopropyl *N*-(3-chlorophenyl) carbamate	(a) Organic solvents (b) Emulsifiable concentrate, wettable powder, granules	Grass killer used by soil application
BCPC	*sec*-Butyl *N*-(3-chlorophenyl) carbamate	(a) Organic solvents (b) Wettable powder	Preemergence herbicide
IPC	Isopropyl *N*-phenylcarbamate	(a) Organic solvents (b) Emulsifiable concentrate, wettable powder, granular form	Grass killer used through the soil; temporary
Eptam (EPTC)	Ethyl-*N*,*N*-di-*n*-propylthiolcarbamate	(a) Organic solvents (b) Emulsifiable concentrate	Soil fumigant at low volatility
Chlorate	Sodium chlorate	(a) Water (b) Dry crystals	General translocated spray, soil sterilant
Cyanamid	Calcium cyanamid, $CaCN_2$	(a) Insoluble (b) Powder and granules	Temporary soil sterilant
KOCN	Potassium cyanate	(a) Water (b) Dry powder	Selective spray against weeds in onions, crabgrass in lawns
OCH (Oktone)	Octochlorocyclohexanone	(a) Oil (b) Concentrate	Fortifier for oil
Endothal	Disodium 3,6-endoxohexahydrophthalate	(a) Water (b) Liquid concentrate	Preemergence and general contact herbicide
HCA	Hexachloroacetone	(a) Oil (b) Liquid	Fortifier for weed oil; contact and soil action

Some important herbicides (Cont.)

Common name or designation	Chemical name	(a) Soluble in: (b) Available as:	Use
MH (maleic hydrazide)	1,2-Dihydropyridazine-3,6-dione, diethanolamine salt	(a) Water (b) Liquid concentrate	Grass killer; spray plus tillage on quack grass
Dichlone	2,3-Dichloro-1-naphthoquinone	(a) Insoluble (b) Dusts and wettable powders	Algicide
NPA (Alanap)	*N*-1-Naphthyl phthalamic acid	(a) Insoluble (b) Granules	Preemergence application
NaNPA (Alanap-3)	Sodium *N*-1-naphthyl phthalamate	(a) Water (b) Water solution and granules	Preemergence spray
Oil, weed	Aromatic bottoms	(a) Oil (b) Liquid	General contact spray crop desiccant
Oil (aromatic solvent)	Xylene type oil fraction	(a) Oil (b) Emulsifiable liquid	General contact agent for aquatic weed control
Oil (selective)	Stoddard solvent	(a) Oil (b) Liquid	Selective light oil for use in carrot crops, forest tree nurseries
DN (general)	4,6-Dinitro-*o*-*sec*-butyl and amyl phenols; 3,5-dinitro cresol	(a) Oils (b) Liquid concentrates	Fortifiers for oil and oil-emulsion general contact sprays
DN (selective)	Salts of dinitrocresol and butyl phenol	(a) Water (b) Water suspension, alcohol solution	Selective sprays for broad-leaved weeds
DN (selective) (DNBP)	Ethanolamine salt of dinitro-*sec*-butyl phenol	(a) Water (b) Liquid concentrate	Preemergence
NaPCP	Sodium pentachlorophenate	(a) Water (b) Dry crystals	Fortifier for oil-emulsion general contact sprays
PCP	Pentachlorophenol	(a) Aromatic oil (b) Concentrate in oil	Fortifier for oil and oil-emulsion general contact sprays
2,4-D	2,4-Dichlorophenoxyacetic acid	(a) Salts in water, esters in oil, acid in organic solvents (b) Water-soluble salts, oil-soluble esters, and emulsifiable acid	Selective translocated spray and preemergence herbicide, broad-leaved weeds
MCPA	2-Methyl-4-chlorophenoxyacetic acid	(a) Salts in water, esters in oil (b) Water-soluble salts; oil-soluble esters	Selective translocated spray; broad-leaved weeds
2,4,5-T	2,4,5-Trichlorophenoxyacetic acid	(a) Organic solvents (b) Water-soluble salts, oil-soluble esters, and emulsifiable concentrates	Selective translocated spray for woody species
4-(2,4-DB)	4-(2,4-Dichlorophenoxy)-butyric acid	(a) Salts in water, esters in oil (b) Salts and esters	Selective translocated spray
4-(MCPB)	4-(2-Methyl-4-chlorophenoxy)-butyric acid	(a) Salts (amine) in water; low volatile esters in oil (b) Salts and esters	Selective translocated spray; special broad-leaved weeds
4-(2,4,5-TB)	4-(2,4,5 Trichlorophenoxy)-butyric acid	(a) Salts in water, esters in oil (b) Salt and ester	Selective translocated spray
2-(2,4-DP)	2,4-Dichlorophenoxypropionic acid	(a) Salts in water, esters in oil, acid in organic solvents (b) Liquid concentrate	Selective translocated spray; oaks and other woody species
MCPP	2-Methyl-4-chlorophenoxypropionic acid	(a) Salts in water, esters in oil, acid in organic solvents (b) Liquid concentrate	Selective translocated spray

Some important herbicides (Cont.)

Common name or designation	Chemical name	(a) Soluble in: (b) Available as:	Use
2,4,5-TP (Silvex)	2,4,5-Trichlorophenoxypropionic acid	(a) Salts in water, esters in oil, acid in organic solvents	Selective translocated spray on woody species
Sesin	2-(2,4-Dichlorophenoxyethyl) benzoate	(a) Organic solvents (b) Emulsifiable concentrate, granules	Preemergence spray, selective
3Y9	Tris-(2,4-dichlorophenoxyethyl) phosphite	(a) Organic solvents (b) Emulsifiable concentrate and granules	Selective pre- or postemergence spray, or dry application
Erbon (Baron)	2-(2,4,5-Trichlorophenoxy)-ethyl-2,2-dichloropropionate	(a) Oil (b) Emulsifiable concentrate	General soil sterilant
Sesone (Crag 1)	Sodium 2,4-dichlorophenoxyethyl sulfate	(a) Water (b) Dry powder	Preemergence spray, selective
2,4,5-TES (Natrin)	Sodium 2,4,5-trichlorophenoxyethyl sulfate	(a) Water (b) Dry powder	Preemergence spray
Fenac	*2,3,6-Phenylacetic acid*	(a) *Water* (b) Liquid concentrate	*Selective against certain* grasses and broad-leaved weeds
Dalapon	Sodium 2,2-dichloropropionate	(a) Water (b) Dry powder	Selective translocated grass killer
F-B_2	1,1′-Ethylene-2,2′-dipyridylium dibromide	(a) Water (b) Liquid	Postemergence general contact spray; desiccant
AMS (ammate)	Ammonium sulfamate	(a) Water (b) Dry crystals	Translocated spray on woody species
DMTT (Mylone)	3,5-Dimethyl-1,3,5,2*H*-thiadiazine-2-thione	(a) Insoluble (b) Wettable powder	General pesticide for soil application; controls many weed species
CDT (Geigy 444)	2-Chloro-4,6-bis(diethylamino)-3-triazine	(a) Organic solvents (b) Emulsifiable concentrate	Selective preemergence spray
Simazin	2-Chloro-4,6 bis(ethylamino)-*sec*-triazine	(a) Insoluble (b) Wettable powder	Selective pre- or postemergence spray
Propazine	2-Chloro-4,6-bis(dipropylamino)-*sec*-triazine	(a) Insoluble (b) Wettable powder	Selective preemergence herbicide
Trietazine	2-Chloro-4-(diethyl-6-ethyl-amino)-*sec*-triazine	(a) Insoluble (b) Wettable powder	Selective preemergence herbicide
Atrazine	2-Chloro-4-(ethyl-6-isopropylamino)-*sec*-triazine	(a) Insoluble (b) Wettable powder	Selective preemergence herbicide
ATA (Amitrol)	3-Amino-1,2,4-triazole	(a) Water (b) 50% soluble powder	Translocated spray on perennial species
DCU (dichloral urea)	1,3-*bis*(2,2,2-Trichloro-1-hydroxyethyl) urea	(a) Insoluble (b) Wettable powder	Preemergence herbicide
Monuron	3-(*p*-Chlorophenyl)-1,1-dimethylurea	(a) Insoluble (b) Wettable powder	Soil sterilant, selective at low dosage, general at high
Diuron	3-(3,4,Dichlorophenyl)-1,1-dimethylurea	(a) Insoluble (b) Wettable powder	Soil sterilant, selective at low dosages, general at high dosage
Neburon	3-(3,4-Dichlorophenyl)-1-methyl-1-*n*-butylurea	(a) Insoluble (b) Wettable powder	Special selectivity at low solubility
Urox	3-(*p*-Chlorophenyl)-1,1-dimethylurea trichloroacetate	(a) Organic solvents (b) Granules and liquid concentrate	General soil sterilant
Herbisan-5	Bis-ethyl xanthogen	(a) Oil (b) Emulsifiable concentrate	Selective preemergence herbicide vegetable and field crops

are sprayed with 2,4-DB or MCPB (the butyric acid analogs of 2,4-D and MCP), they are injured because these phenoxybutyric acids are broken down to the corresponding acetic acids by the enzymes. Crops sprayed with the same solutions are not affected if they lack such enzymes because the chemicals are not broken down and the butyric acids are not toxic.

Many other as yet undisclosed mechanisms of protoplasmic selectivity exist. Undoubtedly the continuing search for new herbicides will uncover chemicals having unique selective properties and these will become the successful weed killers in the future.

Turning now to the six categories in the classification given above, it appears that each involves a group of chemicals having a common mode of action. These will now be discussed in turn.

Selective contact sprays kill weeds in crops as a result of their presence in solution on the foliage. Examples are salts of iron and copper, sulfuric acid, sodium dinitrocresylate, and ammonium dinitro-sec-butyl phenylate. Selectivity is largely a matter of differential wetting. Selective oils used in carrot and other umbellifer crops, cotton, and forest tree nurseries are contact selective materials that wet the crops as well as the weeds. Selectivity is protoplasmic.

Selective translocated spray chemicals move within the sprayed plants. They are useful as selective sprays in low-volume application and as killers of perennial weeds. Examples are the chlorophenoxyacetic acid, propionic acid, and butyric acid, amino triazole, and dalapon.

Selective chemicals applied through the soil include most of the materials that are applied by the preemergence method, and some that are applied postemergence Examples are 2,4-D, MCP, IPC, CIPC, TCA, PCP, chloroacetamides, thiocarbamates, the Crag herbicides, and low dosages of the substituted ureas and symmetrical triazines.

Nonselective contact sprays kill all vegetation to which they are applied. Examples are sodium arsenite, sodium chlorate, penta- and metaborates in solution, ammate, DN compounds in oil, aromatic oils, PCP, and endothal.

Nonselective translocated sprays are used on perennial weeds primarily. Examples are acid arsenicals, chlorates, thiocyanates, and sulfamates.

Materials that cause a general killing of all plants through the soil include fumigants and some of the older chemicals. Examples are carbon disulfide (CS_2), chloropicrin, methyl bromide, EPTC, sulfamates, thiocyanates, arsenates, chlorates, borates, and high dosages of Erbon, the newer, substituted ureas, and symmetrical triazines.

In addition to the six categories mentioned, there are a few recognized method[illegible]
the s[illegible]
water [illegible]
the tops of the weed are bent over and submerged in the solution. As the foliage is injured and rendered permeable, the solution enters the stems and [illegible] and it may pass through interconnected roots to other unsubmerged tops. This method has been used for treating wild morning glory in strawberry beds and garden areas; for eliminating poison oak in camping areas; and for eradicating camel thorn on ranges. Sodium arsenite is the cheapest chemical to use, but its poison hazard must be kept in mind

Killing of aquatic weeds with chlorinated benzene or aromatic solvents is accomplished by pumping the chemical into irrigation water flowing in a ditch.

Properties of herbicides. The table gives some of the important properties and methods of using available herbicides. Both common and chemical names are given. Solubility is expressed only in relative terms. Some chemicals termed insoluble may dissolve in the soil solution in sufficient quantities to be toxic, but they cannot be conveniently formulated or applied in water solution. *Availability* again refers to the forms in which the chemicals are available on the market or are submitted for testing by the manufacturers.

Many of the chemicals listed in the table are used in proprietary mixtures. These are usually marketed under brand names and, if properly formulated, may combine the properties of two or more toxicants. Examples are the borate-chlorate mixtures that combine the quick-killing action of sodium chlorate with the residual action of the borates. They have the additional virtue of safety in contrast to sodium chlorate by itself. The substituted ureas are also used in combination with borates and chlorates. Some chemicals combine two distinct toxicants in a single molecule; examples are Erb[illegible] SCIENCE
(PLANT) [illegible]
PLANT C[illegible]
[illegible]ZATION.

Bibliography: G. H. Ahlgren, G. C. Klingman and D. E. Wolf, *Principles of Weed Control*, 1951; W. W. Robbins, A. S. Crafts, and R. N. Raynor, *Weed Control*, 1942.

Hermaphroditism

A condition in which both gonads—testes and ovaries—are present in the same individual. True hermaphroditism is quite rare and occurs in three general forms: the presence of bisexual gonads, or ovotestes; the presence of both an ovary and a testis; and a combination of an ovotestis with either an ovary or a testis. Usually, the gonad of one sex is not completely functional but a few cases on record have shown function of both sexes.

The internal sexual organs may develop in almost any manner with either a preponderance of those of one sex or a mixture. The external organs may be definitely male or female, but more often show combinations of various degrees.

Pseudohermaphroditism indicates the presence of the gonads of one sex and some degree of development of the internal and external organs of the opposite sex.
[illegible] n the human,
[illegible] om the primi-
[illegible] into either

testes or ovaries normally occurs during the following 2 weeks and depends initially on the type of sex chromosome present. A Y chromosome will induce testicular formation, and an X chromosome, ovarian development. The opposite sex structures become vestigial, although most do not disappear completely. In addition, every person, in a sense, is bisexual in that both male and female hormones are normally produced in various proportions throughout life. A hormonal imbalance, therefore, may stimulate structures of the opposite sex, especially if complete development has not been attained. *See* REPRODUCTIVE SYSTEM DISORDERS.

[E.G.ST.]

Hernia

An abnormal protrusion of a portion of the body through its containing wall. It is sufficiently common to have acquired a popular name, rupture.

Occurrence. Most hernias occur in areas of innate weakness in the body wall. Although such a weakness, and even an area of faulty closure of the wall, may have been present from birth, there is often no obvious bulge until late in life. After a major physical effort or a series of minor strains, various organs may suddenly protrude through the weakened area as a noticeable and often painful lump. Hernias in other instances occur through areas weakened by injury. Most frequently this is surgical incision, but violent stress or infection may also be responsible. The development of a hernia is also favored by conditions which cause general loss of muscle and connective tissue tone. Most striking of these is aging. Dietary factors seem to play an important role among certain semistarved populations. The contents of a hernia vary greatly depending on the mobility of nearby organs and the size of the internal opening or neck of the hernia. The intestines are commonly involved because they normally move about freely in the abdomen. Most abdominal hernias are covered internally by a sac of peritoneum, and externally by stretched-out muscle or fascia, and skin. Some hernias reach enormous size if the basic defect is large, so that all the abdominal organs may slide into the scrotum or chest cavity. Hernias are said to be reducible if the contents can be pushed back into their original location by gentle pressure, incarcerated if the contents can no longer be pushed back but are still healthy, or strangulated if the contents are caught in such a way that their blood supply is cut off. Strangulation creates a surgical emergency. Unless it is relieved there rapidly ensues a leakage of bacteria through the bowel wall into the peritoneal cavity and in turn a spreading peritonitis. Strangulation is favored by narrowness of the neck of the hernia or by swelling of the contents, for example, by gaseous distention of the bowel.

Types. Each of the common types of hernia presents special features. The most common type is the inguinal hernia. The weakness lies in the inguinal canal through which the vas deferens passes from the abdominal cavity to the scrotum; there is an analagous rudimentary structure in the female. The bulge presents above the inguinal ligament and often extends down into the scrotum or labium. Umbilical hernia, through the defect left by the passage of umbilical cord structures during fetal life, is particularly common in infancy, and often closes off spontaneously. *Femoral hernia*, originating in the weakness produced by passage of the major arteries and veins from the abdomen into the legs, presents below the inguinal ligament and is common in women. *Diaphragmatic hernia*, in which a large defect exists in the muscular diaphragm separating chest from abdomen, may be present at birth or may follow a severe injury in later life. In either event the abdominal organs rise into the chest cavity and may seriously impair breathing. Esophageal hiatus hernia is a small bulging of part of the stomach through the diaphragm around the esophagus. It occurs principally in older persons, producing indigestion and anemia. Incisional hernias following major abdominal surgery may be the result of improper closure of the layers of the body wall, or of infection destroying the wall, or of removal of so much tissue that the edges cannot be completely apposed.

There are other less common types of hernia. All present the same noxious features of potential strangulation or of a painful swelling. These features can be completely alleviated only by effective closure of the opening. [R.N.B.]

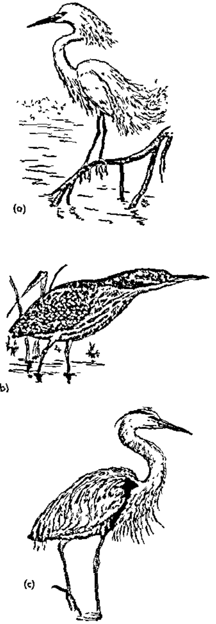
Bibliography: L. M. Zimmerman and B. Anson, *Anatomy and Surgery of Hernia*, 1953.

Heron

Any of several species of wading birds of the family Arteidae. This is a cosmopolitan family of 66 species, 13 of which occur in the United States. Some of them are called egrets and bitterns. They have long, straight, pointed bills, long legs, and long necks. Several of them are large birds, but [illegible] usually nest in trees and in colonies. Bitterns are [illegible] streams.

The egrets are like other herons except that during the breeding season they are adorned with 30–50 long plume feathers called aigrettes which grow from the middle of the back and extend to, or beyond, the end of the tail. By the early 1900s, the egrets were seriously depleted by plume hunters. Although still hunted in some parts of the world, [illegible] September.

The American bittern, Botaurus lentiginosus, is marked with contrasting, irregular longitudi



(a) Snowy egret, Leucophoyx thula; length 24 in. (b) American bittern, Botaurus lentiginosus; length 28 in. (c) Great blue heron, Ardea herodias; length 50 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

stripes, and when alarmed it becomes motionless, with its bill held perpendicular. Its slender build and streaked coloration add to the camouflage, providing it with remarkable concealment. The least bittern, *Ixobrychus exilis*, only 13 in. long, is the smallest of the herons and is a shy, secretive marsh bird with the habits of a rail. *See* RAIL.

Best known of the typical herons is the great blue heron, *Ardea herodias*, a wading bird 48 in. tall, which nests over all of the United States and southern Canada. *See* CICONIIFORMES.

[J.D.B.]

Herpes simplex

A mild vesicular eruption of the skin or mucous membranes caused by a virus. The most familiar form is seen as cold sores or fever blisters.

Primary infections are chiefly inapparent; the commonest clinical form of the first attack is acute herpetic gingivostomatitis, a severe febrile disease with mouth eruptions. Recurrent attacks in persons with antibodies produce local lesions, often at mucocutaneous junctions; attacks frequently follow nonspecific stimuli, such as fever.

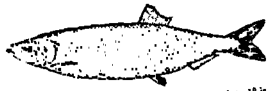
Herpes simplex virus is probably more constantly present in man than is any other virus. Primary infection occurs early; 70–90% of adults have antibodies. However, the virus may be carried throughout life. Transmission is by contamination from saliva or stools. *See* ANIMAL VIRUS.

[J.L.M.]

Bibliography: T. M. Rivers and F. L. Horsfall, Jr. (eds.), *Viral and Rickettsial Infections of Man*, 3d ed., 1959.

Herring

The common herring, *Clupea harengus*, the most important food fish in the world. The herring lives on both sides of the North Atlantic. Over 3,000,000,000 lb are taken annually with an average weight of less than ½ lb. The number of herring alive, if known, would be an astonishing figure; certainly there are more herring than individuals of any other vertebrate species.



The Atlantic herring, Clupea harengus; length to 18 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Herring are marketed in virtually every possible form. The Maine and Norway sardine is the young of this species. Herring travel in great schools in open water, and feed upon plankton. They come inshore to spawn, each female laying about 30,000 eggs.

A closely related form, *Clupea pallasii*, lives in the North Pacific. *See* CLUPEIFORMES. [J.D.B.]

Hesperornis

A genus of fossil, flightless, diving birds of Late Cretaceous age, known from the Niobrara formation of western Kansas; reported also from Montana and South Dakota. The family Hesperornithidae includes *Coniornis*, and, with the Baptornithidae, forms the order Hesperornithiformes. The two best known species, *Hesperornis regalis* and *H. gracilis*, were about 1½ meters long, with strong legs and paddlelike feet placed far back on

the elongated body. The wing was reduced to one slender, curved bone, the humerus, weakly articulated with the shoulder girdle. While perfectly adapted for aquatic life, the angle of articulation of the legs with the pelvis indicates that they could not have stood erect on land. The long jaws had pointed teeth in grooves, a character found only in *Hesperornis* among true birds.

No close relatives of this group are known. Supposed ancestral relationships with ostriches, loons, or grebes appear due to resemblances through convergent evolution. *See* AVES; AVES FOSSILS; NEORNITHES. [A.W.]

Heterobasidiomycetidae

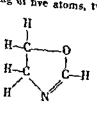
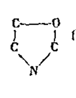
A subclass of basidiomycetous fungi (Basidiomycetes). They are also known as Hemibasidiomycetes or Phragmobasidiomycetes. In the subclass, the basidium is either branched or is divided into cells by cross walls. Some jelly fungi (order Dacromycetales) have branched basidia with the appearance of a tuning fork. Basidia of jelly fungi of the order Tremellales have longitudinal walls. In the remaining orders, which include other jelly fungi, the rusts, and the smuts, the walls are transverse. In most *Heterobasidiomycetidae*, the basidium consists of a first-formed portion, the hypobasidium, and a later-formed epibasidium which intervenes between the hypobasidium and the sterigmata (stalks) upon which basidiospores are borne. *See* BASIDIOMYCETES; RUST (MICROBIOLOGY); SMUT (MICROBIOLOGY). [R.M.P.]

Heterochlorida

An order of the class Phytomastiogophorea. These flagellate organisms are yellow-green to green in color and have numerous chromatophores. Two subequal flagella are usually present. Three suborders, the Euheterochlorina, Rhizochloridina, and Heterocapsina, are recognized by protozoologists. Phycologists classify these groups as orders in the class Xanthophyceae of the division Chrysophyta. *See* ALGAE; PHYTOMASTIGOPHOREA. [C.B.C.]

Heterocyclic compounds

A major class of organic compounds. Organic cyclic compounds are those compounds containing a group of atoms covalently bonded to form a ring. Homocyclic or alicyclic compounds are cyclic compounds in which all the ring atoms are carbon. Heterocyclic compounds are cyclic compounds in which the ring includes at least one atom of an element other than carbon. An example is oxazoline (*I*), which is heterocyclic because it contains a ring of five atoms, two of which are not carbon,


(I)

(II)

as shown in (II). A large proportion of the total number of organic

site. The number of atoms in the heterocyclic ring can range from three to many. There is no a priori limitation to the size of the ring. The ring may contain only single bonds, it may include one or more double bonds, or it may possess the aromatic unsaturation characteristic of benzene. The number of heteroatoms in any one ring is commonly one or two, less commonly three or more. Heterocyclic compounds with more than one ring, either heterocyclic or homocyclic, are not exceptional.

Although the more familiar heteroelements are oxygen, nitrogen, and sulfur, many other elements have been, or could be, incorporated into a heterocyclic system. An a priori requirement is that the

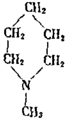
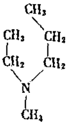
classed as heterocyclic. Metal chelates from di- or polydentate ligands, although heterocyclic, are generally and arbitrarily excluded from this class. Another limitation to the kind of heteroelement is geometric in nature, in that the atom must be able to fit into the ring. The covalence bond angle as well as the bond distance must be compatible with the geometric requirements of the ring when it is considered as a whole.

Natural and synthetic compounds. Heterocyclic systems are encountered in many groups of organic compounds, both synthetic and natural. For example, most alkaloids, sugars, vitamins, and enzymatic cofactors are heterocyclic. Many natural

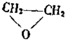
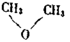
acids are heterocyclic. Actually, the only major groups of natural products that are not heterocyclic are the fats and most, but not all, terpenes, steroids, and protein amino acids. Important synthetic dyes, such as phthalocyanines, cyanines, and phthaleins, as well as many drugs, poisons, and medicinals (both natural and synthetic), for example, sulfathiazole, pyrethrin, rotenone, strychnine, reserpine, certain of the antihistaminics, serotonin, the ergot alkaloids, caffeine, cocaine, morphine, the barbiturates, and the hydantoins, are heterocyclic. Apart from the heterocyclic natural high polymers starch and cellulose, several synthetic polymers are heterocyclic, for example, fur-

tar, are heterocyclic. Many other examples could be cited.

In considering chemical properties, the behavior of saturated heterocyclic rings of five or more members—as a first approximation—resembles that of appropriate open chain analogs. For example, the chemistry of *N*-methylpiperidine (*III*)


(III)

(IV)

is not very *different from* that of methylethylpropylamine (IV). *Often*, differences in behavior of cyclic and *noncyclic analogs* originate in stereochemical factors. One *such factor* concerns four- and three-membered *heterocycles* in which the ring bonds are *distorted with* respect to bond angle and bond length, *just as* they are in cyclobutane and cyclopropane. *These* strained *heterocyclic* systems show enhanced activity in processes *involving* ring opening. Thus, ethylene *oxide* (*V*), although formally an ether, is far *more reactive* than its open-chain analog, dimethyl ether (*VI*). Some polyun-


(V)

(VI)

saturated heterocyclic *systems*, *to a* greater or lesser extent, exhibit chemical properties *implied in* the term aromatic. Pyridine (*VII*), for example, is a thermally stable *material*, resisting addition and oxidation, undergoing *electrophilic* substitution, and possessing *considerable* resonance energy. Thiophene (VIII), *a five*-membered sulfur heterocycle, has the *same set* of aromatic [illegible]

[illegible] ...ctions derived from [illegible]

[illegible] ...ystematic *names* that are preferred over the systematic *names*. One systematic scheme makes use of the *name of* the corresponding alicyclic compound, *and* designates heteroatoms by prefixes such as *oxa* for oxygen, aza for nitrogen, and thia for *sulfur*. According to this method, pyridine (VII) could be called azabenzene, thiophene (*VIII*) would be thiacyclopentadiene, and ethylene oxide (*V*) would be oxacyclopropane. Actually, the *trivial* names for these three systems are commonly *accepted*, and *consequently* the systematic names are not used. Another naming scheme *recommended* for rings of nine or fewer members retains *the* oxa-aza convention for


(VII)

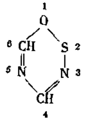
(VIII)

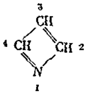
Suffixes for degree of unsaturation

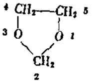
Ring size	With nitrogen: Max	2	1	0	Without nitrogen: Max	2	1	0
3			irine	iridine			irine	irane
4		ete	etine	etidine		ete	etene	etane
5	ole		oline	olidine	ole		olene	olane
6	ine	*	*	*	in	*	*	inane or ane
7	epine	*	*	*	epin	*	*	epane

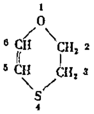
* Add prefix such as dihydro and tetrahydro to the name of the ring with maximum unsaturation.

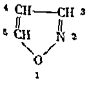
the heteroatoms, but discards the name of the alicyclic compound in favor of a set of defined symbols. The size of the ring is specified by two-letter combinations placed just after designation of the heteroatom(s). The symbols ir, et, ol, in, and ep stand for 3-, 4-, 5-, 6-, and 7-membered rings, respectively. The ending of the name is formed from *suffixes indicating* degree of unsaturation. The *accompanying* table gives various combinations. The *illustrations* given below will show how the systematic naming rules are applied and will also show how the ring atoms are numbered.

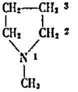

1,2,3,5-Oxathiadiazine


Azete or azacyclobutadiene

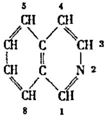

1,3-Dioxolane


2,3-Dihydro-1,4-oxathine


1,2-Oxazole, or preferably, isoxazole



1-Methylazolidine, or preferably, 1-methylpyrrolidine



2-Azanaphthalene, or preferably, isoquinoline

For details about specific heterocyclic systems, *see* ACRIDINE; AZOLE; CARBAZOLE; CHELATION; DIAZINE; FURAN; IMIDAZOLE; INDOLE; ISOQUINOLINE; MORPHOLINE; OXAZOLE; PYRAN; PYRAZOLE; PYRIDINE; PYRIMIDINE; PYRROLE; QUINOLINE; THIAZOLE; THIOPHENE. [W.J.GE.]

Bibliography: American Chemical Society, Editorial reports on nomenclature, *Chem. Abstr.* 39(3):5867, 1945; 41(3):8175–8184, 1947; 46: 12407, 1952; R. C. Elderfield (ed.), *Heterocyclic Compounds*, 6 vols., 1950–1957; A. A. Morton, *The Chemistry of Heterocyclic Compounds*, 1946; A. M. Patterson and L. T. Capell, *The Ring Index*, Am. Chem. Soc. Monograph 84, 1940; E. H. Rodd (ed.), *Chemistry of Carbon Compounds*, vol 4, 1957; A. Weissberger (ed.), *The Chemistry of Heterocyclic Compounds*, 9 vols., 1950–1957.

Heterodyne principle

The basic principle underlying the operation of a superheterodyne radio, television, or other receiver, wherein two alternating currents that differ in frequency are mixed in a nonlinear device to produce two new frequencies, corresponding to the sum and the difference of the input frequencies. Only the difference frequency is commonly used in a superheterodyne receiver, where it serves as the input to the intermediate-frequency amplifier. The heterodyne principle permits conversion of a wide range of different input frequencies to a predetermined, lower intermediate-frequency value that can be amplified more efficiently. Some frequency meters also use the heterodyne principle when comparing an unknown input frequency with a calibrated frequency standard. *See* RADIO RECEIVER. [J.MR.]

Heteronemertini

An order of the class Anopla in the phylum Rhynchocoela. Heteronemertini includes the family Lineidae which contains some of the most common and best-known species which have been used in experimental studies of regeneration and embryology. Among these are *Lineus ruber* and *Cerebratulus lacteus*. Heteronemertines have a body wall composed of three muscular layers: an inner longitudinal layer, a middle circular layer, outer longitudinal layer. A middorsal present and usually there are connections with the lateral vessels. Cerebral organs generally occur as well. *See* ANOPLA; PALAEONEMERTINI. [C.B.C.]

Heterophile antigen

The serologic reactions of the tissue and blood-cell antigens of most animals are normally characteristic of the species, and significant serologic cross reactions are given only with antisera to the corresponding antigens from closely related species. The numerous groups of heterophile antigens—of which the Forssman antigens are the best studied—constitute significant exceptions. Heterophile antigens link the species hog-ox-human (blood group A), cat-horse, dog-hog-cat-human, while several heterophile groups link otherwise diverse microorganisms. Links between pneumococcus type XIV and the human blood groups are also known. The cross reactions between the *Proteus* bacillus and the *Rickettsia* are of importance in the diagnosis of typhus fever. Heterophile antigens need not be present in all organs of an animal and, in particular, if present in the red blood cells, they are likely to be absent in the tissues.

In 1911, J. Forssman reported that the injection of tissues from the guinea pig into rabbits would stimulate the formation of antibodies that, together with complement, would lyse sheep red blood cells. The organs of the horse, dog, cat, mouse, fowl, and tortoise also stimulate the production of antibodies with affinities for sheep cells, while the organs of man, the rabbit, ox, rat, goose, eel, and frog do not. The Forssman antigen is also widely distributed among microorganisms, for example, pneumococci, anthrax, and dysentery bacilli. Although somewhat heterogeneous serologically, the Forssman antigens are all believed to be relatively heat-stable lipopolysaccharides. Forssman antisera, containing antibodies against the Forssman antigen, are employed in the laboratory detection of food adulteration by horse meat, and in the diagnosis of infectious mononucleosis. *See* ANTHRAX; ANTIBODY; ANTIGEN; BACILLARY DYSENTERY; INFECTIOUS MONONUCLEOSIS; PNEUMOCOCCUS; RICKETTSIOSES; SEROLOGY. [H.P.T.]

Bibliography: G. S. Wilson and A. A. Miles, *Topley and Wilson's Principles of Bacteriology and Immunity*, 3 vols. 4th ed., 1955.

Heterophyiasis

Presence in the small intestine of man of the minute intestinal fluke *Heterophyes heterophyes*, 1.5 mm long. The Nile Delta and Orient are endemic foci. *See* TREMATODA.

Many reservoir hosts exist. Eggs in feces are ingested by brackish-water snails, Potamididae, and cercariae finally emerge from these to enc in the flesh of fish, *Mugil* and *Tilapia*. of the raw fish flesh by man completes *See* PELECYPODA; PISCES (ZOOLOGY). bers of the worm damage the mucosa. diarrhea, colicky pains, and ggs may enter lymph or blood to

brain, with serious complications. Tetrachloroethylene is the indicated curative drug. *See* PARASITOLOGY, MEDICAL. [J.F.MA.]

Heterosis

Hybrid vigor or increase in size, yield, and performance found in hybrids, especially if the parents have previously been inbred. The application of heterosis has been one of the most important contributions of genetics to scientific agriculture in providing hybrid corn and, more recently, vigorous, high-yielding hybrids in other plants and in livestock. *See* BREEDING (ANIMAL); BREEDING (PLANT); GENETICS.

It has been known for several centuries that hybrids between varieties and even between species are frequently of unusual size and vigor. The proverbial hardiness of the mule is often given as an example in animals. Charles Darwin devoted many years to a study of the effects of inbreeding and hybridization in several kinds of domestic plants. He reported, and his findings have since been abundantly confirmed, that inbreeding inevitably leads to decreased size and weakness, and to an increased frequency of abnormalities. The changes are cumulative so that with successive generations of inbreeding the deleterious effects are more and more pronounced. However, all the effects of inbreeding are immediately eliminated by outcrossing. Darwin also noted that many plants have elaborate mechanisms that prevent self-fertilization, thus avoiding the deleterious effects of close inbreeding.

A satisfactory explanation of heterosis awaited the development of Mendelian genetics. It is now known that the effect of inbreeding in a population is to make homozygous many genes that had previously been heterozygous. Therefore the decline in vigor and size must have its explanation in the increased homozygosity that accompanies inbreeding.

Likewise, heterosis may be explained by the reverse effect of increased heterozygosity following outcrossing, since a hybrid will in general be more heterozygous than its parents. *See* MENDELISM.

Hypotheses. There are two principal hypotheses to account for the association of size and vigor with heterozygosity.

Dominance. The dominance hypothesis notes that any noninbred population carries a number of recessive genes that are harmful to a greater or lesser extent, but which are rendered ineffective by their dominant alleles. As they become homozygous through inbreeding, they exert their harmful effect. With hybridization, some of the detrimental recessives contributed to the hybrid by one parent are masked by dominant alleles from the other, and an increase in vigor is the result. This hypothesis is supported by numerous experimental studies showing the large number of harmful recessives carried in actual populations, a number that is quantitatively adequate to account for the observed decline in vigor with inbreeding and recovery on crossing.

Overdominance. The alternative hypothesis is that there are loci at which the heterozygote is superior to either homozygote in vigor. This, the overdominance hypothesis, also has the conse[illegible] that vigor is proportional to heterozygosity. [illegible] erty, but on the other hand such genes would be expected to persist in a population and to make a disproportionate contribution to its variability. The dominance hypothesis has been more widely ac[illegible] particularly in determining why one hybrid is better than another.

Use of heterosis. The most important example of the systematic use of heterosis has been the development of high-yielding hybrid corn. Almost all the corn grown in the north central corn-producing states of the United States is now hybrid. In the years around 1910 two American geneticists, E. M. East and G. H. Shull, discovered that by inbreeding corn and then making hybrids between the inbred strains they could obtain excellent plants. The hybrids not only had a high yield but also were extremely uniform, and by choosing the right inbred strains for hybridization the breeder could incorporate into the hybrids other desirable traits such as disease resistance, straight stalks, and well-shaped ears.

The main difficulty in this scheme is that the seeds from which the hybrid plants develop are grown on ears of inbred plants. Hence the seed producer has a low yield, and the seed becomes expensive. This difficulty is circumvented by using the double cross, or 4-way cross, originated by D. F. Jones. The breeder starts with four inbred strains and makes two different crosses yielding two hybrids. The two hybrids are then crossed to produce the commercial double-cross seed. In this way, the commercial seed is produced on high-yielding hybrid plants. The double-cross hybrids have about the same yield as the single-cross hybrids though they are somewhat less uniform.

Following the pattern set for hybrid corn, breeders have produced many other agricultural plants as hybrids. There has also been an increasing tendency to use heterosis in animal breeding, both by employing mating schemes that maximize heterosis and by hybridization between inbred lines, the latter being widely practiced especially in poultry breeding. [J.F.C.]

Bibliography: J. W. Gowen (ed.), *Heterosis*, 1952.

Heterostraci

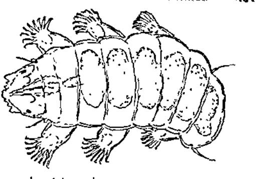
An extinct group of ostracoderms or armored, jawless vertebrates. The armor has a distinctive microscopic structure, consisting of bone lacking any cavities for bone cells, surmounted by tubercles or ridges of dentine. Fragments of such armor from the Middle Ordovician of Colorado and Wyoming

represent the earliest known vertebrates. Heterostraci became more common towards the end of the Silurian and persisted through the Devonian.

The anterior part of the body was covered with plates, which were few in number in Silurian forms but became more numerous in later genera. The posterior part of the body and the tail were covered with thick scales, sometimes very large in Silurian genera. With the exception of the tail, there were no fins. The eyes were far apart at the sides of the head, and the paired nostrils probably lay just inside the mouth. There were no jaws, but the mouth was bounded behind by a number of small plates that may have been used for nibbling. There were numerous gills that had a single external opening on each side. Well known Heterostraci are the Devonian *Pteraspis*, one species of which attained a length of 5 ft, and *Drepanaspis*, which had a broad, flat body adapted for life on the sea bottom. See AGNATHA; OSTRACODERM. [R.H.DE.]

Heterotardigrada

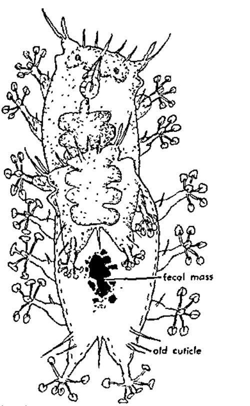
An order of the tardigrades, the majority of whose genera have widely varied structure. Cephalic appendages having a sensorial function are present, as well as cirrus lateralis and clava. Pharyngeal pockets are strengthened by uninterrupted ridges, except in *Pseudechiniscus islandicus*, which has interrupted ones. Toes or claws are uniform in structure and completely separated from one another. A preanal gonopore is present; excretory glands are absent. This order of tardigrades is divided into two suborders, Arthrotardigrada and Echiniscoidea.



Echiniscoides sigismundi.

Arthrotardigrada is a suborder of heterotardigrades whose members have toelike terminations of the legs. The tubular middle part of the leg telescopes into the broad proximal part. These animals are marine organisms found in sand or on algae. One species, *Tetrakentron*, is found on the buccal tentacles of a sea cucumber, *Leptosynapta*. *Batil-*

retractable into the proximal part. At least the fourth pair of legs has a distinct fold. There are two families: Nudechiniscidae, with a uniform cuticle; and Scutechiniscidae, with segmental and intersegmental thickenings (plates) of cuticle. Frequently these animals are red in color due to the presence of carotenoid pigments. Many species are active at relatively low humidity. See TARDIGRADA. [E.M.]

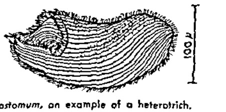


Batillipes, defecation during molt.

Heterotrichida

A large order of the Spirotricha containing many well-known, sizable species. The buccal ciliature, both membranes and the adoral zone of membranelles, is well developed, although in a number of families the somatic, or body, ciliature is really holotrichous in nature. Heterotrichs have become adapted to all sorts of habitats including the digestive tracts of a variety of invertebrate and a few vertebrate hosts. Man is not involved. Some species contain pigments in their cytoplasm giving them such coloration as blue, green, pink, brown, or black.

Because of their size and amazing regenerative powers, a number of heterotrichs have been widely used in experimental research paralleling the studies by embryologists in the fields of morphogenesis and differentiation. *Stentor* (most popular in morphogenetic investigations), *Blepharisma*,



Climacostomum, an example of a heterotrich.

... species,

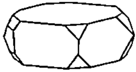
),

of

3 mm ($\frac{1}{8}$ in.), are common species. *Nyctotherus* is found in the digestive tract of amphibians and many invertebrates. *Balantidium*, long considered a heterotrich and important as the only ciliate parasitizing man, is actually a trichostome. *See* SPIROTRICHA; TRICHOSTOMATIDA. [J.O.C.]

Heulandite

A mineral belonging to the zeolite family of silicates and crystallizing in the monoclinic system. It usually occurs in crystals with prominent side pinacoid, often having a diamond shape. There is perfect side pinacoid cleavage on which the luster is pearly; elsewhere the luster is vitreous. The crystals often have undulating faces, and are made up of subindividuals in nearly parallel position. In polarized light they show optical anomalies of a



Crystal with prominent side pinacoid, typical of mineral heulandite. (From C. S. Hurlbut, Jr., *Dana's Manual of Mineralogy*, 16th ed., Wiley, 1952)

sectoral nature. The hardness is 3½–4 on Mohs scale; specific gravity is 2.18–2.20. The mineral is usually white or colorless but may be yellow or red. *See* ZEOLITE.

Heulandite is essentially a hydrous calcium aluminum silicate, $Ca(Al_2Si_7O_{18})\cdot 6H_2O$. Small amounts of sodium and potassium usually substitute for calcium. Heulandite is a secondary mineral found in cavities in basalts associated with other zeolites and calcite. Notable localities are in the Faeroe Islands, India, Nova Scotia, and West Paterson, New Jersey. [C.FR.; C.S.HU.]

Hexactinellida

A class of the phylum Porifera which includes sponges with a skeleton made up basically of hexactinal siliceous spicules. Spongin is absent. Hexactinellida usually are radially symmetrical in form and have cylindrical, vase-, funnel-, or trumpet-shaped bodies. Some species are fan-shaped and others exhibit a dendritic or labyrinthine mode of branching. Some species attach to a hard substratum directly by a basal attachment region, but many have basal tufts or mats of spicules which serve to anchor them in bottom sediments. The Hexactinellida are exclusively marine and occur in the deeper waters of all seas. Their bathymetric range extends from depths of about 25 down to 8500 meters, but very few descend into the hadal zone, the faunal zone of the deep trenches.

Morphology. The anatomy of the soft parts of Hexactinellida differs from that of the other two

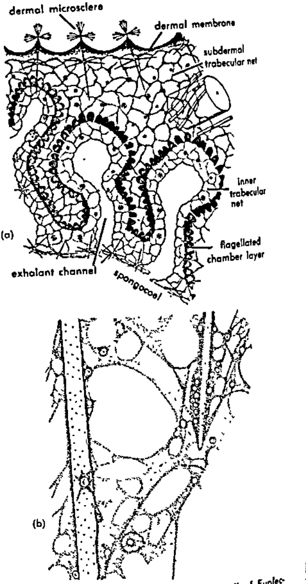


Fig. 1. (a) Section through the body wall of *Euplectella* (from Hyman, 1940, after Schulze, 1887). (b) Portion of the outer trabecular layer of a young specimen of *Farrea*, highly magnified (from Hyman, 1940, after Okada, 1928).

classes of sponges. The body wall has a cavernous structure penetrated by intercommunicating cavities across which stretch the soft parts in the form of filmlike trabeculae. Thimbleshaped flagellated chambers, arranged side by side in a single layer, are supported in the midst of the trabeculae. Five regions may be recognized in the body wall. The outer dermal membrane is often filmlike and is perforated by more or less round pores which serve for the entrance of water into the intertrabecular lacunae. The subdermal trabecular network, which is continuous with the dermal membrane, is dense peripherally, but the lacunae become more spacious toward the interior and form subdermal cavities. These are in contact with the outermost flagellated chambers. Fairly distinct inhalant

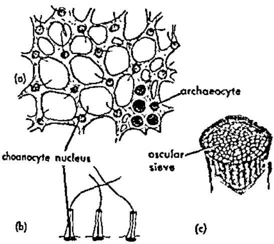


Fig. 2. (a) Reticular membrane of a flagellated chamber of *Euplectella*, highly magnified (after Ijima, 1901). (b) Lateral view of choanocytes of *Euplectella*, showing collars and flagella arising above the nuclei of the reticular membrane, highly magnified (after Ijima, 1901). (c) Upper end of *Euplectella*, showing oscular sieve (after Hyman, 1940).

canals may lead to deeper-lying flagellated chambers in species with thick walls and a folded chamber layer. The flagellated chambers are composed of choanocytes whose bases are continuous and form a reticular membrane perforated with holes, the prosopyles, which allow passage of water from the subdermal trabecular lacunae into the flagellated chambers. Each chamber opens widely through an apopyle into the subspongocoelic trabecular system which in turn opens into the central spongocoel by way of a perforated membrane. Fairly distinct exhalant channels are often present in the trabecular system spreading between the chamber layer and the spongocoel. The spongocoel opens through a large terminal osculum, covered over in some genera by a sieve plate, the meshwork of which is reinforced by spicules.

Although the trabeculae and flagellated chambers are apparently syncytial in nature, free archaeocytes and their derivatives, thesocytes and reproductive cells, are found on the trabecular strands. Mesoglea is absent, at least in postlarval stages.

Skeleton. The skeleton is made up of megascleres and microscleres. Although many forms of spicules occur, most of them can apparently be derived from the triaxon or hexactin, in which three axes meet at right angles. Through loss of rays, pentacts, tetracts, triacts, and diactinal and monactinal spicules are formed. Some monactinal types, however, give no evidence of derivation from the triaxonid form. The megascleres are also classified according to their position in the body of the sponge. Three basic categories occur in the body wall: dermal, those lying in or just beneath the dermal membrane; parenchymal, those lying in the interior trabecular net, chiefly outside the chamber layer; and gastral, those lying in or beneath the membrane bounding the spongocoel. Dermal and gastral megascleres almost always lie free and unconnected to one another in the soft tissue, but parenchymal megascleres often become fused to their neighbors during the development of the sponge. Megascleres which protrude from the body wall are called prostals and may form a marginal fringe around the osculum, protrude from the sides of the body, or form a basal tuft or mat.

Aside from those which are simply small versions of megascleres, the microscleres are of two main types, hexasters and birotulates. The former are small hexactinal spicules which usually have branched ends; the latter are small monaxons, diaxons, and triaxons in which the distal end of each ray bears an umbrellalike expansion.

The spicules of Hexactinellida are laid down in syncytial scleroblast masses A thin film of protoplasm persists around the spicules, at least at the ends of the rays, after their formation. The spicules are composed of hydrated silicon dioxide laid down around an organic axial thread in concentric layers, said to be separated by thin films of organic material.

Development. Developmental stages are known in only a few of the Hexactinellida. It is not known whether sperm is transferred to the egg by a carrier cell as in other sponges. In *Farrea* cleavage results in a solid blastula. Later the embryo consists of an outer layer of closely packed cells and a central mass of transparent jellylike material in

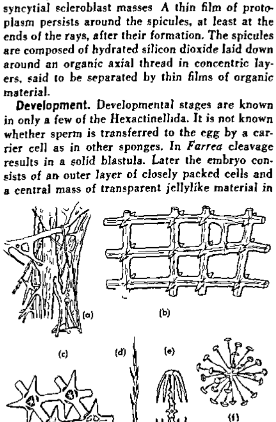


Fig. 3. (a) Portion of lyssacine skeletal framework of *Euplectella*. (b) Portion of dictyonine skeletal framework of *Farrea*. (c) Portion of lychniscose framework of *Aulocystis* (after Reid, 1958). (d) Dermal pentact of *Hyalonema*. (e) Amphidisc of *Hyalonema*. (f) Discohexaster of *Dictyocalyx*. (g) Floricome of Rec (after Schulze, 1887).

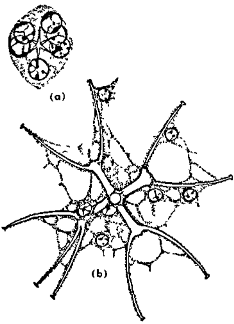


Fig. 4. (a) Early stage in secretion of discohexaster of *Farrea* (b) Later stage of same showing 4 rays. (*From Hyman, 1940, after Okada, 1928*)

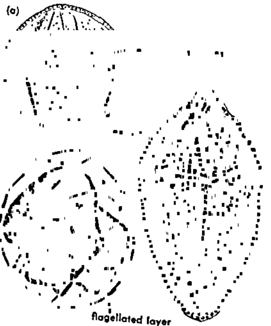


Fig. 5. (a) Early larval stage of *Farrea*. (b) Later stage of same. (c) Young attached sponge of *Farrea*. (*From Hyman, 1940, after Okada, 1928*)

which amoeboid cells wander. Cross-shaped tetractinal spicules appear early in development. When the larva leaves the adult sponge there are 12 of these spicules, peripherally situated. Flagellated chambers appear before the larva escapes, but the collars and flagella apparently form after the larva is set free. After attachment the sponge assumes a tubular form with a terminal osculum. Pentacts soon develop, followed by hexactinal and other spicule types.

Ecology. Other animals often live in association with Hexactinellida. The proximal part of the root tuft of *Hyalonema* is often encrusted with zoanthids or pedunculate barnacles. Sea anemones and hydroids live on some sponges of this class, and isopods and brittle stars are known to inhabit the cavernous interior regions. Several species of *Euplectella* harbor shrimps (*Spongicola*) in their spongocoels. Usually a pair is found, although sometimes only one occurs. Because of the rigidity of the skeletal network in *Euplectella*, the shrimps are imprisoned for life.

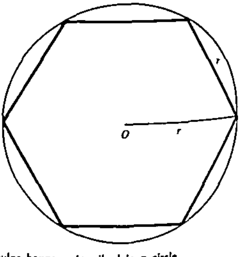
Hexactinellida are well known as fossils because their rigid skeletons, often composed of a lattice of fused spicules, increase the chances of their being preserved intact. Sponges of this class are reported from the Cambrian period upward. There is a notable peak of abundance of fossil remains in the Cretaceous period. *See* Porifera. [W.D.H.]

Hexactinosa

An order of the subclass Hexasterophora in the class Hexactinellida. The parenchymal megascleres in this order of sponges are united to form a rigid framework. The parenchymal megascleres consist wholly of simple hexactins which are arranged in parallel linear series, the members of each series being united one to another by a secondary envelope of silica. Examples include *Hexactinella*, *Aphrocallistes*, *Eurete*, and *Farrea*. *See* Hexactinellida; Hexasterophora. [W.D.H.]

Hexagon

A closed figure formed by the six line segments, or sides, that join in order six ordered points, or vertices, of a plane. In elementary geometry it is usually assumed that the sides of a hexagon do not cross and even that the figure bounds a convex region of the plane. The ancient Greek geometers investigated regular hexagons, that is, hexagons with all sides equal and all angles included by adjacent sides equal 120°. They found that a regular hexagon is obtained by applying the radius of a circle six times as a chord to the circumference.

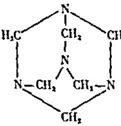


Regular hexagon inscribed in a circle.

and that three regular hexagons completely fill up the plane about a point. For Pascal's theorem concerning hexagons *see* CONIC SECTION. *See also* POLYGON; POLYTOPES, REGULAR. [L.M.BL.]

Hexamethylenetetramine

A white, odorless, water-soluble compound having a cage structure of four 6-membered rings of three carbons and three nitrogens, $(CH_2)_6N_4$. Formal-



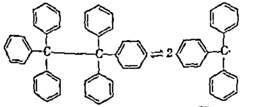
dehyde and ammonia condense to hexamethylenetetramine and water. Slow reversibility of this reaction allows its use as a source of either ammonia or formaldehyde, the latter as disinfectant or urinary antiseptic. Nitration in the presence of ammonium nitrate and acetic anhydride gives very efficient conversion to trimethylenetrinitramine (*cyclonite*, hexogen, RDX), an explosive of higher brisance than trinitro toluene (TNT). It was manufactured for use with TNT in large bombs during World War II. Alkyl halides slowly ammonolyze in acid to primary amines in the presence of the compound (Delepine reaction). *See* EXPLOSION AND EXPLOSIVE; FORMALDEHYDE. [L.B.C.]

Hexaphenylethane

A colorless, crystalline hydrocarbon (I) which melts with decomposition at 145–147°C. Moses Gomberg (1866–1943) first prepared the hydrocarbon from triphenylchloromethane by the action of silver or zinc or mercury metal in the absence of air.

If hexaphenylethane, protected from air or oxygen, is dissolved in benzene or carbon disulfide, it gives a yellow solution. Colorless hexaphenylethane may be obtained again by evaporation of the solvent. If a solution of hexaphenylethane is exposed to air, it becomes colorless, yielding a peroxide $(C_6H_5)_3COOC(C_6H_5)_3$, and if it is treated with iodine, it yields triphenylmethyl iodide.

Gomberg correctly concluded that the hydrocarbon (I) dissociates in solution yielding triphenylmethyl radicals (II). The extent of the dissociation is dependent upon concentration and is greatest at higher dilutions. Dissociation also increases with



(I) (II)

higher temperature. The degree of dissociation in solution is measured by boiling-point elevation, freezing-point lowering, or colorimetric methods, or by measurement of paramagnetic susceptibility. The last method is possible because free radicals have an unpaired electron and are attracted to a magnet; the undissociated hexaphenylethane is not.

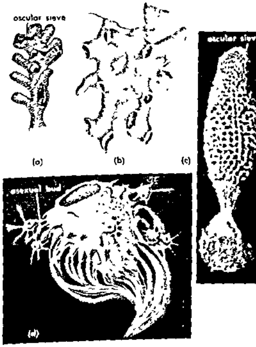
Other hexaarylethanes which are even more highly dissociated than hexaphenylethane have been synthesized. For example, hexa-(*p*-biphenylyl)-ethane is 100% dissociated in 2–3% benzene solution at 5°C while under the same conditions hexaphenylethane is only 1–3% dissociated into radicals. *See* ALKYL RADICAL; FREE RADICAL; TRIPHENYLMETHANE. [C.K.B.]

Hexapoda

A class of the phylum Arthropoda used frequently as a synonym for Insecta. The term Hexapoda was proposed in 1825 by P. Latreille. The three pairs of appendages, specialized as legs, are characteristic features of the hexapods. The hexapods are considered to include two main groups by some systematists, the wingless forms or the Apterygota, and winged forms or the Pterygota. Other authorities separate the Protrua from those groups considered to be the true insects. *See* INSECTA. [C.B.C.]

Hexasterophora

A subclass of sponges of the class Hexactinellida, in which the parenchymal microscleres are typically hexasters (small, six-rayed spicules, often



Representative hexasterophorans. *(a) Aphrocalli (after Hyman, 1940).* (b) *Eurete (after Schulze,* (c) *Polylophus* showing asexual buds (after 1887). (d) *Euplectella oweni* (after Ijima, 190

with branched ends). This is a diverse assemblage of sponges commonly firmly fixed to the substratum by the base, less commonly anchored by means of basal spicule tufts or mats. The spicules of the body are sometimes free and unconnected, but the parenchymal megascleres are often fused to form a rigidly connected skeleton. The following orders are recognized: Hexactinosa, Lychniscosa, Lyssacinosa, and Reticulosa. *See* HEXACTINELLIDA; HEXACTINOSA; LYCHNISCOSA; LYSSACINOSA; RETICULOSA. [W.D.H.]

Hexode

A six-electrode vacuum tube, consisting of a cathode, four grids, and an anode. It is generally used as a mixer tube in superheterodyne receiver circuits. The nonlinear characteristics of the tube are used in such a way that when a radio-frequency signal is applied to one grid and a signal from a local oscillator is applied to another grid, the beat-, or difference-, frequency component appears in the plate current. Thus the mixer tube functions as a frequency converter.

The electrodes of the hexode are operated in the following way: the cathode is at zero voltage and acts as the source of electron current; the first grid is operated at a small negative potential and serves to inject the local-oscillator voltage; the second grid is operated at a large positive voltage and acts as a screen grid to reduce the electrostatic coupling between the signal and oscillator grid; the third grid is operated at a small negative voltage and serves to inject the radio-frequency signal voltage; the fourth grid is operated at a large positive voltage and serves as a screen to reduce the electrostatic coupling between the signal and output circuits; the plate is operated at a larger positive voltage and acts as the collector of the modulated electron current.

When the hexode is used as a mixer tube, its operation is described in terms of a parameter known as the conversion transconductance. This is defined as the ratio of the current output at the intermediate frequency to the signal-voltage input at the radio frequency. Under best operating conditions the conversion transconductance will be approximately one-third of the transconductance between the grid into which the radio frequency is injected and the plate. The hexode was one of the first mixer-type tubes developed. Since then other types have been made which exhibit better operating characteristics. Hexodes in general suffer from some interaction between the two input circuits, a relatively low conversion transconductance, and a relatively low plate resistance. *See* VACUUM TUBE. [K.R.S.]

Hibernation

A general term applied to a condition of dormancy and torpor found in cold-blooded vertebrates and invertebrates. It is difficult to define for warm-blooded vertebrates because it does not refer to a single specific condition in these animals.

COLD-BLOODED VERTEBRATES AND INVERTEBRATES

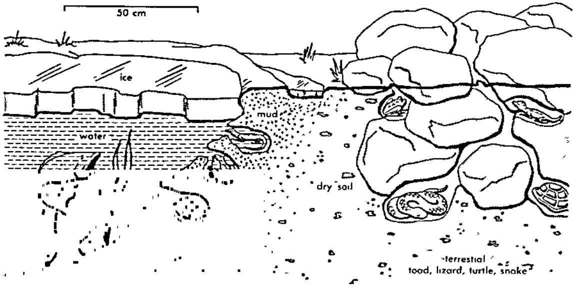
Hibernation occurs with exposure to low temperatures and, under normal conditions, occurs principally during winter seasons when there are lengthy periods of low environmental temperatures. Animals that hibernate naturally can be induced to do so under controlled laboratory conditions; consequently, hibernating animals are becoming useful as experimental subjects. A related form of dormancy is estivation. Many animals estivate when exposed to prolonged periods of drought or during hot, dry summers. For all practical purposes hiber-

either a cold or an arid environment.

There is no complete list of animals that hibernate; however, many examples can be found among the poikilotherms, the cold-blooded animals, both vertebrate and invertebrate. The poikilotherms are sometimes referred to as ectothermic, because their body temperatures are not internally regulated but, rather, follow the rise and fall of environmental temperatures. During hibernation, body temperatures reflect the environmental temperature, often within a fraction of a degree. Among the classic examples of hibernators or estivators are reptiles, amphibians, and fishes among the vertebrates, and insects, mollusks, and many other invertebrates.

Hibernating reptiles. Many terrestrial reptiles such as lizards, snakes, and turtles become dormant and hibernate by burrowing in crevices under rocks, logs, and in the ground below the frost line. Terraqueous turtles also become cold-torpid and may often be found completely submerged in mud and in ponds under ice.

Since the hibernating reptile is subject to the caprices of duration of seasonal low temperature, there is no well-defined period of dormancy. The period of hibernation may often be related to latitudinal positions, as evidenced by the turtle family Emydae. These are aquatic and semiaquatic species, some of which range in the United States from northern latitudes southward through the Mississippi Valley. Species that inhabit the northern climes will hibernate longer than their southern relatives, thus showing hibernation periods which are proportional to the length of the winter period. The physiological characteristics of these hibernating reptiles show many striking alterations of body chemistry. Hibernating reptiles show a loss of appetite and discontinue the ingestion of food. Although the metabolic rate is reduced as much as 95% in hibernating turtles, there is some utilization of stored food products. There are two principal types of reserve food: (1) lipids, which are often found in solid fatty masses in the viscera, under the skin, and distributed in tissues such as muscle, kidney, and, particularly, liver; and (2) glycogen, the animal starch, which is less stable and more rapidly used than fats. Glycogen is generally localized in tissues such as liver and muscle. There is evidence that these reserve foods are selectively



Hibernacula of various cold-blooded vertebrates.

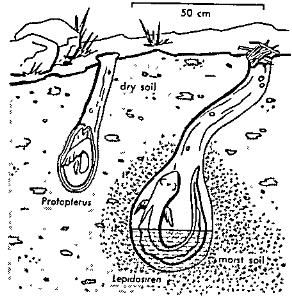
utilized. In hibernating turtles, the tissue glycogen is used during the initial days and weeks of hibernation; later, the lipids are utilized.

The Crocodilia, such as caimans, crocodiles, and alligators, are least able to hibernate. These live in tropical regions where the winter temperatures are relatively uniform. The American alligator, *Alligator mississippiensis*, becomes lethargic and stops eating for several months during winter. This animal undergoes other seasonal physiological modifications as is evidenced by a winter hypoglycemia, that is, a reduction in blood sugar level. The blood sugar level of animals is a physiological index of metabolism and hormonal balance. A major hazard to hibernating poikilotherms is death from freezing; ice crystals form in free protoplasmic water and ultimately destroy the cells and tissues, causing the death of the animal. Frogs, salamanders, and turtles are often surrounded by ice while in hibernation. These animals are able to survive, despite the reduction in body temperatures to about 0° to −1°C. As winter approaches, the water content of the tissues becomes reduced and the blood more concentrated. Experiments with turtles have shown that, during winter, there is a lowering of the freezing point of blood. Insects are also credited with the ability to lose body water during winter. Therefore, any physical adjustments which these animals are able to make to rid themselves of water will aid their survival. Hibernating frogs die if, during a severe winter, the frost penetrates their mud hibernacula. Frequently, the uppermost layers of mud freeze and so do its inhabitants, while the animals buried further down in the unfrozen mud remain safe. Frogs, as well as fishes, often freeze to death in shallow artificial pools made of concrete and lacking a mud substrate.

In consideration of the geographical distribution of the amphibians, the limitations imposed by permafrost and the frozen tundra of the northern latitudes can readily be realized. Amphibians are not ordinarily found in permafrost climates.

Fishes. Hibernation in fishes is not common. Many fishes do, however, spend much of the winter in a state of quiescence while partially frozen in mud and ice. Shallow ponds in temperate, subarctic, and Arctic latitudes often freeze solid in winter. Fishes such as the carp may become encased in the ice or partially buried in the bottom mud. Thus, they remain torpid for lengthy periods. *Dallia pectoralis*, the Alaskan black fish, has received some fame regarding its ability to "freeze solid." Actually, *Dallia*, or any other fish, does not freeze, but becomes encased in ice. If the fish were to freeze solid, with the production of ice crystals in its tissues, it would die. Fishes exposed to cold have a much-reduced metabolism; therefore their oxygen requirements are readily adjusted to the oxygen content of the cold waters. The situation is similar to that which occurs in hibernating aquatic turtles.

The phenomenon of estivation is best known in the dipnoans, that is, the lungfishes. These fishes, *Neoceratodus* in Australia, *Protopterus* in Equatorial Africa, and *Lepidosiren* in South America, are restricted to tropical regions marked by repeated seasons of drought. They survive the dry seasons by becoming dormant and torpid. The lungfishes and their relatives were abundant during the Paleozoic era. They are, therefore, among the more primitive air-breathing animals possessing a lung which utilizes atmospheric oxygen. This lung becomes the primary organ of respiration during the torpidity of estivation. In general, the lungfishes follow a similar behavioral pattern as the dry seasons approach. *Protopterus*, for example, burrows in the bottom mud as the water begins to diminish during the dry season. Once in its burrow, the fish assumes the coiled posture of seasonal dormancy. A lifeline of air is provided by the tunnel



Lungfishes, *Protopterus* and *Lepidosiren*, in estivation.

the burrow to the surface. In preparation for estivation, *Protopterus* secretes a slimy mucus around itself which hardens in a tight cocoonlike chamber. This chamber forms an effective hygroscopic structure which prevents the desiccation of the fish. There is but one opening, formed around the mouth. Thus the air from the tunnel enters the mouth and passes to the lung apparatus. At the termination of the dry season, water slowly enters the burrow, softens the contents, and awakens the lungfish. The metabolism of the lungfish is at a low ebb during estivation, with the energy for its modest life processes provided by the utilization of tissue protein. This is evidenced by the large increase in the formation of urea, a product of protein metabolism. The urea is promptly excreted in large quantities when the fish returns to water. *Lepidosiren*, the South American member of the family of lungfishes, sometimes called the mud-siren, also burrows. The burrow entrance is loosely plugged with surface clay, perforated with air holes, each about 0.5 in. in diameter. Mud-sirens burrow in the deepest and most concave bottoms, seemingly to ensure that they will receive the inflowing water when it first returns to the pond. The fishes' tunnel-like burrows are about 2–3 in. in diameter and extend downward about 2–3 ft. The mud-siren estivates in a gourd-shaped, gelatinous-coated burrow which may be partially filled with water. These fishes, therefore, differ from their African relatives in that they maintain a soft, moist medium. Lungfishes in estivation are readily handled and transported, indicating the extremely quiescent nature of their torpor.

Hibernation and estivation in invertebrates. This is often more than a period of seasonal dormancy, because in some snails estivation may be extended for years at a time, and among the insects and spiders the period of hibernation becomes intimately associated with a phase in the life cycle. During the winter months and during a hot dry summer, the soil contains a remarkable variety of torpid invertebrates, for example, earthworms, snails and slugs, nematodes, insects and spiders, grubs, larvae and pupae of many insects, egg cases and cocoons. The hibernacula for dormant invertebrates are also provided by rotting logs and vegetation, the undersurfaces of bark, and crevices in rock and building foundations.

Insecta. Insects overwinter, for the most part, in the egg or larval stage of metamorphosis (egg→larva→pupa→adult). Hibernation frequently becomes integrated with the diapause, or arrested development, of the egg or larva which occurs during the winter. The egg of the silkworm, *Bombyx mori*, undergoes diapause in conjunction with hibernation. The eggs are laid in autumn and must be exposed to about 0°C for several months to ensure hatching in the spring. The Syrphidae, commonly referred to as the hover flies, are ordinary garden inhabitants. They live in close association with plants, as commensals and parasites, and are also known to infest animals, producing myiasis (see MYIASIS). The larvae of the syrphids are particularly susceptible to hibernation and estivation. They become markedly desiccated but are quickly reactivated by exposure to water. The Lepidoptera, butterflies and moths, offer some of the most common examples of hibernating larvae and pupae. The familiar cocoon of the butterfly is the hibernaculum of the larva and pupa. The disappearance of mosquitoes in the Arctic, subarctic, and temperate zones in autumn and winter is also a reflection of insect hibernation. The adult hibernates, after its last autumnal blood meal, and presumably subsists on its stored food reserves. Provision for overwintering in *Culex* and *Anopheles* mosquitoes lies in their development of a reserve food material in the form of a fat body. *See* INSECT PHYSIOLOGY.

Mollusca. Land mollusks, such as snails and slugs, are noted for their ability to hibernate and estivate. The land snails are particularly cold-hardy and have frequently been found encased in ice. Among the more cold-resistant land snails are species of *Physa*, which normally live in northern Siberia, 73°30′N latitude where the mean annual temperature is below −20°C. In the subarctic and temperate zones most of the fresh-water mollusks can live in water under the ice and are often enclosed in solid ice. The common garden snail, *Helix aspersa*, and the edible snail, *Helix pomatia*, exhibit the chief characteristics of hibernation common to the Gastropoda. In autumn, these snails cease ingesting food and burrow into the ground under logs, leaves, and other forms of natural debris, where they remain in hibernation for about 6 months. The animal makes this burrow with its muscular foot and settles down to a winter sleep. The hibernating position is such that the shell aperture is turned upward. The hibernating snail may be found singly or in clusters, with numerous

individuals in contact with each other. The animal seals its aperture by secreting a membrane, the epiphragma, chiefly composed of calcium phosphate. In the epiphragma, there remains a minute respiratory opening. The respiratory movements are drastically reduced and, as in other poikilotherms, the rate of heart beat is related to temperature; at 30°C the heart rate is 50–60 beats per minute, whereas in hibernation the rate is reduced to about 4–12 per minute. The energy for the lowered metabolism during hibernation appears to be derived from the utilization of stored fats, with a subsequent loss of body weight.

The slugs, referred to as garden snails without shells, are also hibernants. Their hibernaculum is similar to that of the shelled forms. They hibernate in the ground under leaves, tree bark, and other forms of decomposing vegetation. Estivation of the land mollusks occurs chiefly in tropical climates, where they bury themselves deeply in the ground or under rocks. They also form a type of epiphragma as a defense against evaporation and desiccation. Tropical snails are able to estivate for lengthy periods, months and, possibly, years at a time. A specimen of *Helix desertorum* is reputed to have awakened after 5 years of dormancy. See MOLLUSCA.

The marine mollusks, periwinkles (*Littorina*) and mussels (*Mytilus* and *Modiolus*), cannot be regarded as hibernators, yet they are extremely cold-hardy. These animals are inhabitants of the intertidal zone. In such latitudes as Woods Hole, Mass., and Hebron, Labrador, they may be exposed repeatedly to temperatures of −20°C in winter. At such temperatures, ice forms inside the shelled animals. These seashore mollusks can live through the winter season with as much as 70–75% of the body cavity being recurrently filled with ice. At such low temperatures the animals are being dehydrated since a great deal of the body water is being converted into unusable ice and the remaining fluid becomes a more concentrated salt solution. In these animals, the body water which freezes is not tissue or cellular protoplasmic fluid, but is chiefly extracellular water.

Nematoda. Nematodes are cosmopolitan and have been found in practically all ecological settings. However, the terrestrial and fresh-water forms are most apt to respond to winter exposure by becoming quiescent. Pasture puddles, which often freeze solid in winter, harbor a particularly rich nematode fauna. Literally millions per acre are found in the top few inches of soil. Among the "cushion" plants such as mosses and lichens one finds frequent nematode "parasitism." In addition to a rich fauna of nematodes, these plants also harbor tardigrades, rotifers, and even protozoans. These animals are characterized by their capacity to endure the extremes of drought, heat, desiccation, and cold. Under such forms of environmental stress they go into a dormant state, and revive when environmental conditions become less severe and again conducive to active life. The particular nematodes such as *Mononchus, Dorylaimus, Plectus, Monhystera, Cephalobus, Trilobus, Trypyla, Tylenchus, Aphelenchus*, and others are considerably alike throughout the world, and in spite of wide longitudinal and latitudinal separation their overwintering and drought behavior are similar.

Their resistance notwithstanding, nematodes do die and decline in numbers during the winter. See NEMATODA; ROTIFERA; TARDIGRADA.

Bryozoa. The Ctenostomes are the only Bryozoa with fresh-water representatives. These animals bear a resemblance to moss and seaweeds and are collectively called moss animals. The colonial fresh-water *Paludicella* and *Victorella* form hibernacula. These are specialized external buds which persist through the winter while the rest of the colony breaks down. At the termination of winter, the shell of the hibernaculum splits into two halves and a young colony is brought forth. See BRYOZOA.

Encystment. The phenomenon of encystment is commonplace in the Protozoa, or single-celled animals. Encystment is remarkably similar to estivation and hibernation, and an encysted protozoan is extremely quiescent and almost nonmetabolizing. It is generally accepted that encystment is a protective measure against cold and desiccation, and yet it may coincide with some general physiological, protoplasmic reorganization or dedifferentiation in the species. For example, during encystment of the ameba *Pelomyxa carolinensis*, there is a reduction in number of nuclei, a concentration of protoplasm, and the formation of a cyst membrane and shell. The ciliate *Colpoda cucullus* during its encystment undergoes nuclear reorganization, and multiplication of the organism takes place. The ability of protozoans to encyst is at least one of the reasons for their world-wide distribution. The Arctic tundra, which remains frozen for 6–8 months of the year, teems with protozoans as well as other invertebrate life. These microscopic animals could not live through these winter rigors unless they had this capacity for self-preservation and protection.

The free-living protozoans offer many examples of encystment due to winter low temperatures and to evaporation due to drought. The cyst may be regarded as the hibernaculum, and it is usually found in mud or ice and soil. Cysts of the single-celled organisms can be recovered from ice and mud samples taken in midwinter and also from dried soil collected during the most arid summers. The protozoan cyst is often, as with many invertebrates, a function of reproduction and therefore a stage in the life cycle of the species. The ciliates *Colpoda* and *Didinium* are noted for their cyst formation. Encysted *Colpoda* can resist experimental freezing for as long as 2 months, and can be activated after being in a dried state for as long as 5 years. See PROTOZOA.

Hibernacula. The hibernacula of poikilotherm vertebrates and invertebrates are as varied as the animals themselves. The minute cyst in protozoan

the cocoon and egg case of insects and spiders, the burrows and crevices of reptiles, and the dried mucous case of the lungfish, in all instances, protect the animal from evaporation or desiccation and freezing. [X.J.M.]

WARM-BLOODED VERTEBRATES

Many mammals and some birds spend at least part of the winter in hiding, but remain no more drowsy than in normal sleep. On the other hand, some mammals undergo a profound decrease in metabolic rate and physiological function during the winter, with a body temperature near 0°C. This condition, sometimes known as deep hibernation, is the only state in which the warm-blooded vertebrate, with its complex mechanisms for temperature control, abandons its warm-blooded state and chills to the temperature of the environment. Between the drowsy condition and deep hibernation are gradations about which little is known. The bear, skunk, raccoon, and badger are animals which become drowsy in winter. Although the bear is usually considered the typical hibernator, measurements have shown that its body temperature does not drop more than 4C°. Body temperature, measured rectally, was 35.5°C at an air temperature of 4.4°C. At −3.5°C, body temperature was 31.2°C.

Deep hibernators. The deep hibernators are confined to four orders of mammals: the Chiroptera or bats, the insectivores, the rodents, and, probably, the primates. Most, if not all, of the insect-eating bats of temperate climates not only hibernate in the winter, but also drop their body temperature every time they roost and sleep. The advantage of this for a small mammal with a disproportionately large heat-losing surface is obvious when conservation of energy is considered. Among the insectivores, the European hedgehog is a deep hibernator, and the Madagascan tenrec (*Centetes*) probably hibernates or estivates during the summer. Many rodents are deep hibernators, including ground squirrels, woodchucks, dormice, and hamsters. The only known primate hibernators are the fat-tailed lemurs of Madagascar which apparently estivate during the cool dry season. Among the birds, the poor-will (*Phalaenoptilus*) and some hummingbirds undergo a marked lowering of body temperature and metabolic rate in cold periods.

With all deep hibernators except the bats, hibernation is seasonal, usually occurring during the cold winter months. In all cases, it occurs in animals which would face extremely difficult conditions if they had to remain active and search for food. A few desert-living species disappear during the driest, hottest months, and it is probable that they enter into a state of deep hibernation in their cool burrows, though no studies have been made on this, or on the estivation of animals during the tropical rainy season. Hibernators are restricted typically to more temperate zones, but the Alaskan ground squirrel lives near the Arctic Circle.

Physiology of hibernation. During a preparation period for hibernation, the animals either become fat like the woodchuck, or store food in their winter quarters, like the chipmunk and hamster. Prior to hibernation, there is a general involution of the endocrine glands, but at least part of this occurs soon after the breeding season and is not directly concerned with hibernation. Animals such as ground squirrels become more torpid during the fall, even when kept in a warm environment, indicating a profound metabolic change which may be controlled by the endocrine glands. On the other hand, ablation of any of these glands has not brought on hibernation. Lack of food will cause hibernation in the spiny pocket mouse (*Perognathus*), but in most hibernators this has little, if any, effect and the stimulus for hibernation is not known.

Hibernation in mammals is not caused by an inability to remain warm when exposed to cold, for most hibernators are capable of very high metabolic rates. In the woodchuck, heart rate and oxygen consumption decline before body temperature as the animal enters hibernation, indicating that low environmental temperature does not initiate hibernation. As hibernation deepens, the heart rate, [illegible] slowly

[illegible]

cur, raising the body temperature temporarily and causing a stepwise entrance into hibernation.

In deep hibernation, the body temperature is only 1–2C° above the environment, and it is a peculiarity of hibernators that the vital processes can function at lower temperatures than those of nonhibernators. The heart rate varies between 3 and 15 beats/minute, and the metabolic rate is one-thirtieth or less of the warm-blooded rate at rest. During hibernation the source of energy is evidently fat, for the respiratory quotient (RQ) is 0.7. The average RQ of fats has been experimentally determined to be 0.707. Blood sugar levels may be reduced in some mammals, but, in general, the hibernator maintains a remarkably rigid control of its internal environment. If, for example, the ambient temperature is decreased to below the freezing point, the hibernator usually increases its metabolic rate and thus maintains its body temperature above the lethal level. If the animal is exposed to an atmosphere containing about 5% carbon dioxide, it increases its respiratory rate as it would if it were active and warm. Therefore, the hibernator is not at the mercy [illegible] -rte- [illegible] is capable of waking at any time, using self-generated heat, and this characteristic clearly separates the hibernating state from any condition of induced hypothermia. During the total period of hibernation, the hibernator spontaneously wakes from time to time, usually at least once a week. In the period of wakefulness the stored food is evidently eaten, but animals which do not store food rely on their fat for the extra energy during the whole winter.

Arousal is a complex, coordinated event which results in warming in the shortest time, usually

about 3 hours when in an environmental temperature of 5°C. As soon as the animal starts to rouse, the heart speeds and oxygen consumption rises, followed shortly by an increase in body temperature. The anterior parts of the body warm rapidly, but because of vasoconstriction, the posterior parts remain cold until late in the warming process. Shivering can be detected early, using the electromyogram, and it becomes more obvious as the animal warms. The energy for warming, at least in rodents, is obtained from glycogen stored in the liver and other tissues. It has been calculated that one waking uses as much energy as does 10 days of hibernation, so that waking must use most of the winter stores, whether fat or fodder.

The precise sequence of events in entering and waking from hibernation and the control of the internal environment while in hibernation suggest that the whole process is mediated by the central nervous system; details are not known. [C.P.L.]

Bibliography: C. P. Lyman and P. O. Chatfield, Physiology of hibernation in mammals, *Physiol. Revs.*, 35:403–425, 1955.

Hickory

Any species of the genus *Carya*, formerly known botanically as *Hicoria*. Hickories are mostly tall forest trees characterized by strong, terminal, scaly winter buds, pinnately compound leaves, solid pith (not chambered), and fruit with an outer husk or exocarp which splits more or less readily into four parts, revealing a nut with a hard shell or endocarp.

(a) Shagbark hickory, *Carya ovata* (b) Mockernut hickory, *Carya tomentosa* (c) Pignut hickory, *Carya glabra*. (d) Bitternut hickory, *Carya cordiformis*. (A. H. Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

The shagbark hickory, *C. ovata*, grows to a height of about 120 ft and is found in the eastern half of the United States and adjacent Canada. It is the most important species because both its nuts, the hickory nuts of commerce, and its wood are of especial value. It is easily recognized by the bark which in older trees exfoliates from the trunk in long, curving, irregular plates, and by the leaves which usually have five leaflets, of which the terminal one is larger.

The pecan, *C. illinoensis*, is also a valuable species because of its commercially popular, thin-shelled, sweet nuts. Although its native range is limited to the Mississippi Valley region and Mexico, many varieties are cultivated in the southern United States.

Other species are the mockernut and pignut hickories, widely distributed through the eastern half of the United States, and the shellbark hickory found in the east-central United States. The remarkably tough and strong wood of all species is the world's best wood for tool handles. It is also used for parts of vehicles, furniture, flooring, boxes, and crates, and for smoking meats. See FOREST AND FORESTRY; JUGLANDALES; TREE. [A.H.G.]

High fidelity

A term (often shortened to hi-fi) applied to sound reproduction, used to designate a state of performance in which realism is achieved. To achieve realism in sound reproduction, four fundamental conditions must be satisfied, as follows: (1) the frequency range must include without frequency discrimination all the audible components of the various sounds to be reproduced; (2) the volume range must permit noiseless and distortionless reproduction of the entire range of intensity associated with the sounds; (3) the reverberation characteristics of the original sound must be approximated in the reproduced sound; and (4) the spatial sound pattern of the original sound must be preserved in the reproduced sound. See ARCHITECTURAL ACOUSTICS; DISK RECORDING; MAGNETIC RECORDING; REVERBERATION; SOUND REPRODUCTION SYSTEMS, ELECTRICAL; STEREOPHONIC SOUND. [H.F.O.]

High-pressure phenomena

Of those variables which have the largest effect on the free energy or chemical activity of any system (defined as a portion of the universe isolated for study) the most important are composition, temperature, and pressure. The vast majority of relevant research in chemistry and physics has been concerned with variations in the first two parameters. There has been a good reason for this; it has been experimentally difficult to change the pressure enough to make an appreciable difference on

a system. It can easily be calculated that to cause the equivalent change in an average, or typical, substance it is possible either to cool it by 1°C or to subject it to a pressure of 100 atmospheres (atm). (A bar equals 0.987 atm. and can be considered to be roughly equal to 15 psi.) Every science laboratory equipped with a bunsen burner can heat a substance to about 1000°C; to achieve the same effect with pressure one would need 100,000 atm, or the equivalent of close to 1,500,000 pounds per square inch (psi). An orientation in the magnitude of high pressures is important. Atmospheric pressures on or above the surface of the earth range from 0 to 14.7 psi. Certainly 99.99% of the chemical reactions studied have been studied at or just below atmospheric pressure. However, high pressures are not uncommon. They range from the 2–5 atm in automobile tires and pressure cookers, through 100 or so atm in the boilers of ships and power plants, to perhaps 100,000 atm at the point of impact of a high-speed rifle bullet and a hard wall High pressures up to 100,000 atm can be generated nowadays in a very simple apparatus using automotive truck jacks. In the laboratories of chemists and earth scientists the pressures and temperatures attainable for studying chemical reactions have increased from about 10,000 atm in 1900 to 25,000 atm in 1940 and to 100,000 atm in 1960. *See* HIGH-PRESSURE PHYSICS.

Areas of high-pressure research. The earth itself is a giant laboratory in which pressures up to about 3,000,000 atm are generated with increasing depth; moreover, in the earth the temperature and pressure increase together (Fig. 1). It is generally agreed, however, that of the rocks actually observed on the surface of the earth, none has been subjected to pressure greater than 10–100,000 atm. A possible exception may be some meteorites which could have originated in the interior of a large disintegrated planet. *See* METEORITE.

The interest of the earth scientist, especially the geochemist, in high-pressure inorganic reactions is therefore natural, since in attempting to understand natural processes leading to the formation of various rocks, it is essential to duplicate the conditions existing in the earth. The science of interpretive petrology rests today on the accumulation of data obtained at high temperatures, high pressures, or both. Theories concerning the composition and properties of the different layers deep within the earth which may explain seismic discontinuities, the magnetism of the earth, possible slippage of one layer over another, and other natural phenomena can be checked only by research of high-temperature and high-pressure processes and reactions.

An additional impetus to research in this field is the ability to produce, under pressure, previously unsynthesized minerals, as well as quite new materials which may be expected to be especially dense and hard. Other phases which contain volatile components, such as water or carbon dioxide, can only be prepared under high pressures of these volatiles. Finally, there is considerable evidence that fluids such as water are excellent catalysts for many inorganic reactions occurring under high pressure.

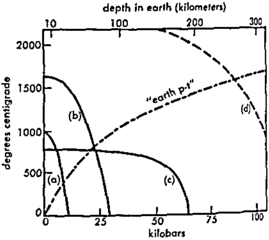


Fig. 1. Experimental limits of various types of apparatus and a generalized pressure-temperature relation of a portion of the earth. (*a*) Simple, externally heated test-tube or cold seal vessels. (*b*) Internally heated, hydrostatic pressure vessels. (*c*) Externally heated, uniaxial pressure devices. (*d*) Internally heated, piston and cylinder devices.

Apparatus in high-pressure research. Much of the recent significant research in this field has been made possible by new apparatus. This in turn owes its origin in large measure to the new materials, such as hard alloys and carbides, produced by an advancing technology. Below are described three or four major families of apparatus with which the majority of high-pressure chemical research is carried on today. In each of these a sample is subjected to fluid or mechanical pressure for a period of time, from minutes up to days or months, at some desired temperature. The reaction is stopped abruptly, or quenched, by rapid cooling and more or less rapid removal of pressure. The resultant products, often in a metastable condition in the laboratory, are examined by using x-ray techniques, the petrographic microscope, differential thermal analysis, infrared absorption spectroscopy, and other more common or more specialized techniques.

The starting materials used in these reactions are often of great importance. The use of amorphous materials such as glasses and gels or other metastable phases is often a decisive factor in the chemical kinetics problem encountered, since time is usually an important limitation with respect to the strength of the equipment.

The types of apparatus commonly used today can be divided into four categories, depending on whether or not the pressure is hydrostatic and

whether the heating is external or internal. The range of pressure and temperature which each of these types can cover is shown in Fig. 1. Also shown are the estimated pressure and temperature within a portion of the earth. The complexity and ease of operation are not inversely related to the pressures attainable, both the externally heated devices being quite simple.

In Fig. 2 more detail is given on the first type, the cold seal test-tube pressure vessel which is the workhorse of much high-pressure research today in the range up to 5000 atm and 1000°C. This extremely simple device can be used to react a very wide range of materials, whether solid or liquid. The materials, whether simple silicates or concentrated HF or NaOH solutions, are sealed into platinum or gold capsules which transmit the hydrostatic pressure actually supplied by an inert fluid outside the capsule. These vessels are heated by an external furnace as illustrated. A related device, which also uses fluid pressure, has the furnace inside the vessel, which is cooled by flowing water. Such devices are ultimately limited by the fact that virtually all gases are frozen at room temperature when the pressure reaches 25,000–30,000 atm. The pressure-temperature (p-t) working range of these types of vessels can be compared by making reference to curves a and b respectively in Fig. 1.

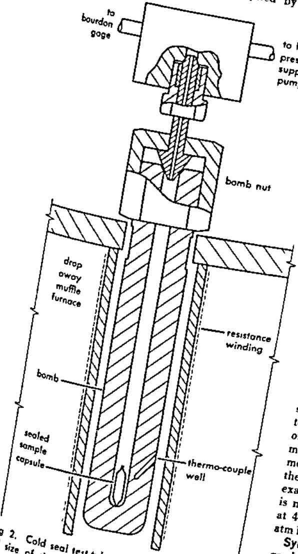


Fig 2. Cold seal test-tube pressure vessel. Basic over-all size of the test tube is 8 in. in length by 1 in. diameter.

The second and simpler type has been evolved from designs by P. W. Bridgman, a pioneer in high-pressure research. As shown in Fig. 3, it uses a hydraulic ram to apply a directed or uniaxial pressure on a small wafer of sample surrounded by a nickel ring and platinum foil pressed between appropriate small-area piston faces. Pressures up to 60,000 bars are possible with this setup, heated externally to temperatures approaching 650°C or slightly higher (conditions schematically shown in c of Fig. 1), depending on the material and design of the pistons. Complex mechanical buttressing of the pistons has been used in some modifications to raise the operating pressures to almost 200,000 atm.

The third type, the prototype of which was first described fully by L. Coes, is shown in Fig. 4. It is essentially an internally heated chamber of a piston and cylinder arrangement with the cylinder walls buttressed and cooled to support internal pressures of 60,000 atm at sample temperatures up to 2000°C. A number of variants of this type of apparatus as well as of the second type are being tried in the effort to reach higher pressures. Some use support from four pistons at tetrahedral angles; others use different ways of supplying the supporting pressure. At this time little systematic work has been reported on experiments with such apparatus, but it is apparent that temperatures of about 2000°C can be sustained for at least a few minutes at a pressure of 100,000 atm.

Another type of apparatus uses the energy of a shock wave to produce the high experimental pressures. Explosive charges or the impact of a rapidly moving column of gas is utilized to produce pressures of the order of 400,000 atm. The coincident temperatures and pressures are difficult to measure or calculate. The duration of these conditions is in microseconds, much too short to allow rearrangements of atoms or ions in crystalline structures and therefore of less interest to the geochemist. For example, although coesite, the dense form of silica, is made quite readily in hours at 20,000 atm and at 400–500°C, shock pressures of 200,000–400,000 atm have failed to accomplish this.

Synthesis of minerals and new materials. A good measure of the success of the earth scientist in the high-pressure field is the fact that since the synthesis of diamonds, it can now be claimed that any material which has been observed in nature can be made in the laboratory. To be sure, there

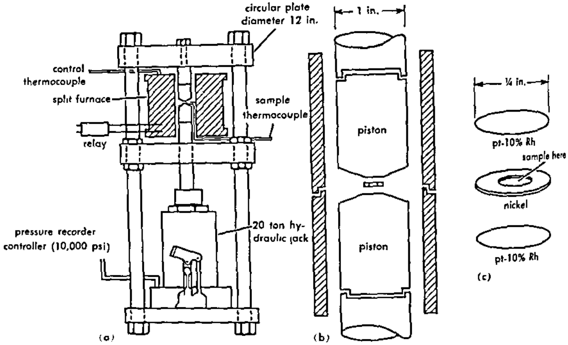


Fig 3 (a) Schematic layout of uniaxial high-pressure apparatus with accessories. (b) Details of high-pressure pistons (c) Sample assembly.

are many which have not been tried, but there is no reason to doubt that in every case the laboratory can duplicate the efforts of nature. *See* DIAMOND; GEM, MANUFACTURED.

First, one may mention the synthesis of the hard, dense phases which have presumably been thrust up rapidly from great depth. The most spectacular of the recent successes is the synthesis of diamonds in a few laboratories. Perhaps more important to the mineralogist has been the synthesis of the minerals jadeite, lawsonite, sillimanite, kyanite, and garnets such as pyrope, almandine, and andradite. All these phases were first synthesized in a dramatic breakthrough by Coes, using the apparatus shown in Fig. 4. Next is the extensive and

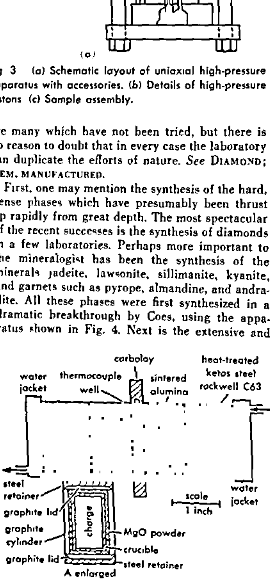


Fig. 4. Coes-type high-temperature–high-pressure apparatus.

systematic work on volatile-containing phases such as micas, clays, and complex carbonates. Under high water pressures all pure or end member phases have been prepared and their properties studied and defined. Moreover, the extent of the systematic replacement of one ion by another has been studied. Isotopic fractionation effects in minerals have likewise been examined in the laboratory, using deuterium for hydrogen and O^{18} for O^{16}.

Many of the most interesting syntheses achieved have yielded phases which do not occur in nature and have never before been prepared. In some cases a long-expected phase was finally prepared, such as the hard cubic form of BN, analogous to diamond. In others, unexpected phases were found; thus, silica, SiO_2, itself the most abundant substance on earth, gives a new form called coesite, some 10% denser than the usual form, quartz. New high-pressure forms of a large number of substances have been found. They cover the periodic table from B_2O_3 and BeF_2 through $MnPO_4$, MnF_2, and $FePO_4$ to PbO_2 and U_3O_4. It must be remembered that these changes include only those transformations which involve such a major rearrangement of strong bonds that, when the high pressure is removed rapidly, the reverse reaction is prevented and the high-pressure form is preserved metastable to ambient conditions. So far only simple compounds of monovalent cations have failed to yield new phases of this type.

Melting under volatile pressure. It is well known that the addition of one substance to another lowers the freezing point of the latter. This is true of the vast majority of substances; the addition of salt to

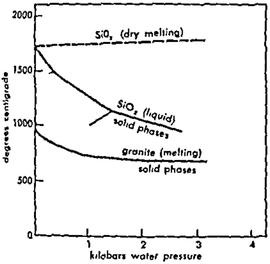


Fig. 5. *Effect of water pressure on melting of silica, SiO_2, and granite, with water entering the liquids. The melting behavior of SiO_2 under inert pressure is shown for comparison.*

water is one of the commonest in experience. For many decades it had intrigued geologists to consider what influence water or other volatile materials would have on the melting behavior of rocks. Naturally, to carry out such an experiment the entire assembly of rock and volatile addition would have to be confined under pressure at or near the melting points of the rocks which are of the order of 1000–1500°C. In Fig. 5 two examples are shown of the influence of water pressure on common rocks and minerals. The lowering of the melting point of SiO_2 from 1730 to 1250°C under only 1000 atm of water pressure is quite dramatic. The water in this case dissolves in the siliceous liquid. Also shown for comparison is the influence of inert or dry pressure on the melting point of silica. In this case the gas or solid transmitting the pressure does not dissolve in either the liquid or solid silica. The similar dry-pressure curve is also shown later for the important mineral diopside, to illustrate the fact that in general the melting point is raised slowly with pressure. The distinction between melting in an inert atmosphere and one in which the volatile phase under pressure dissolves in the liquid should be borne in mind. In deeply buried rocks it can be assumed that much of the water is effectively sealed in the system and may enter the liquid phase. The results of the study of melting the assemblage of minerals which form granite are also shown in Fig. 5. The fact that a granite can be melted as low as 660°C at a pressure of 4000 atm (corresponding to a burial of about 10 miles) with only a few per cent of water in the liquid is one of the most important clues on the origin of many of the commonest igneous rocks in the earth.

Applications in petrology. While the synthesis of new or important compounds has a spectacular aspect, much of the recent effort of physical chemists working with high pressures is concerned only incidentally with synthesis. Instead, the main drive is toward the obtaining of new data on the pressure, temperature, and composition conditions under which certain mineral assemblages are stable. Nature provides several typical assemblages of minerals as characteristic of certain rock types or families. As the conditions under which each particular assemblage is stable are determined, a partial reconstruction can be provided of the conditions which must have existed at any particular place on the earth.

Two different types of reactions may be studied in this connection. The first type involves reactions such as decarbonation or dehydration. Two simple reactions of this type are used in the pressure-temperature diagram of Fig. 6, to illustrate the principle of the application of such data. The general form of such curves is seen to be convex toward the high-temperature and low-pressure side, with very steep slopes at pressures above a few thousand atmospheres, with an asymptotic approach to the temperature axis at very low pressures. On the low-temperature side of such curves, the hydrates such as mica, or carbonates such as magnesite, are stable. Thus, if the pressure from the depth of burial can be estimated by studying the minerals present, it is possible to determine whether or not the rock has been heated to temperatures above the curve. Thus it is possible to explain why the dark micas occur in some high-temperature extrusive rocks (lavas) while the white or muscovite micas do not. The dehydration-decomposition curve for the latter, not shown in Fig. 6, lies some 350°C lower than the dark-mica curve shown, while the temperatures of these rocks are intermediate.

In Fig 7 is shown a compilation on one diagram of several *p-t* curves for assorted yet common compositions Each curve represents a transformation reaction, where the composition of the condensed phases on either side of the *p-t* curve is the same. It may be noted that these curves are quite different from those of Fig 6. They are all essentially straight lines in the pressure range shown. The

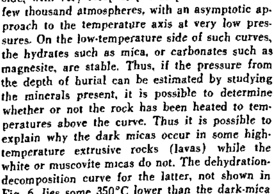


Fig. 6. Typical stability curves for minerals which lo CO_2 or H_2O on heating.

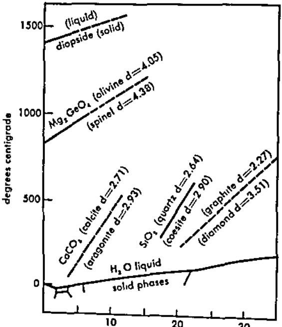


Fig. 7. Curves showing pressure-temperature dependence of typical solid-solid and solid-liquid transformations for single substances.

dense form is favored by high pressure. The pressure theoretically required to make diamonds is actually quite modest. Transformations that require twice the pressure to give the dense forms have been studied quantitatively.

In general, Fig. 7 illustrates how solid-liquid transformations such as the melting of ice do not differ basically from solid-solid transitions. These curves may of course be used in a manner similar to that described for Fig. 6 to suggest whether or not a particular rock was exposed to a certain set of p-t conditions. In practice, curves of both types are used to refine as far as possible the petrologic predictions [illegible] phase (from less dense to more dense forms) in those minerals which geologists believe make up most of the mantle of the earth, it is possible to explain some of the variations in the velocities of seismic pulses at different depths in the earth. The present results do in fact suggest the possibility that the layering in the earth may not reflect changes in composition at all. Indeed, refined and sophisticated measurements of color changes in important minerals such as olivine at pressures up to 150,000 atm indicate that such a mineral may become metallic at pressures of less than 1,000,000 atm. In such phases the neutron and proton cores may eventually be stripped of their extra nuclear electrons to become a sort of metallic plasma. This may well represent the real conditions at the core of the earth. See EARTH INTERIOR. [F.DA.; R.R.]

Bibliography: R. Roy and O. F. Tuttle, Investigations under hydrothermal conditions, in L. H. Ahrens, K. Rankama, and S. K. Runcorn (eds.), *Physics and Chemistry of the Earth*, vol. 1, 1956.

High-pressure physics

High-pressure physics is concerned in general with the alteration in the properties of matter when exposed to the action of high hydrostatic pressures, that is, high pressures exerted uniformly and perpendicularly to all surfaces, as distinguished from nonhydrostatic stresses with shearing components. The properties of matter depend on temperature as well as pressure, but when one speaks of high-pressure physics, the primary interest is in the pressure, and temperature is considered as a parameter modifying the effect of pressure.

The connotation of the term high has undergone a historical evolution. A hundred years ago, when the continuous transition between gas and liquid was of primary interest, the pressures of the critical state, which in general are of the order of a few hundred atmospheres, were regarded as high. Today the critical phenomena are not generally treated as being in the high-pressure range, which may be regarded somewhat arbitrarily as extending from a lower limit of a few hundred atmospheres upward to the present experimental limit. This limit is of the order of a few hundred thousand atmospheres for static prolonged pressures, and a few million atmospheres for the transient pressures attained in shock waves explosively produced.

Under high pressures, the distinction between gas and liquid disappears, and there exists instead a single amorphous state of matter. Under sufficiently high pressure, this single amorphous state may become as hard and viscous as ordinary glass. In addition to the amorphous state, matter under high pressure is capable of existing in the crystalline or true solid condition. Experimentally, a great many liquids can be made to crystallize with discontinuous decrease of volume by the action of sufficiently high pressures. Ideally, as discussed later, this may be true for all liquids, but in practice (for example, under geological conditions), the ideal crystallization may be suppressed by the enormous increase of viscosity produced by pressure in the amorphous phase. In any event, in all high-pressure experimentation when pressure is raised sufficiently high, matter eventually becomes so stiff that pressures are no longer transmitted hydrostatically. In ordinary practice, this may mean pressures of the general order of magnitude of tens of thousands of atmospheres. At these pressures, part of the experimental problem becomes to ensure that the nonhydrostatic or shearing components of stress are as small as possible compared with the hydrostatic component. The transmitting medium now has to be a soft solid, chosen empirically as one which itself does not become prohibitively stiff under pressure. The higher the temperature, the greater the pressure range accessible before stiffening of the pressure medium becomes important.

Experimentation problems. A number of problems of technique had to be solved before high

pressure experimentation could be widely practiced. The problem of preventing leak of the pressure-transmitting medium was the most immediate and obvious of these. A great many methods have now been worked out, valid for different materials and different ranges of pressure. Many of these methods are such that the pressure in the packing is automatically maintained continuously higher than the incident pressure by the action of the incident pressure itself, so that leaks cannot occur, and the possible pressure is limited only by the strength of the containing vessel. It is the containing vessel that sets the present limit. In the early days, the material of the vessel was often defective, and leaks were frequent through ostensibly solid material. This difficulty has now been met, and experimenters are limited only by the intrinsic strength of the material of the containers. Various devices are employed to raise the limits of strength; of these, the most effective are those which utilize the action of the pressure itself to give support to crucial parts of the apparatus. Materials of the highest strength (Carboloy, alloy steels) are needed.

Another problem of technique is that of measurement. There is a fundamental difficulty here because any measuring apparatus is itself distorted by the pressure. It was a long time before the problem of correcting for the distortion of the measuring apparatus was satisfactorily solved. In the lower ranges, pressure is most effectively measured by some sort of dead-weight free-piston gage. Techniques have been worked out for getting electrically insulated leads inside the pressure apparatus, and electrical measurement is used in all present pressure experimenting in determining the effects of pressure on the many physical properties of materials. Many of the techniques possible at low pressures become impossible when pressures become so high as to freeze the transmitting medium. Indirect methods of inferior accuracy, both for measuring pressure and for determining its effects, must be employed.

High-pressure effects. The major effects of high pressure on matter include diminution of volume, changes of physical form, changes of electrical conductivity, and increases in viscosity.

Diminution of volume. The simplest of all the effects of pressure is to diminish volume. The diminution of volume may be thought of as the result of two effects, a diminution in the free space separating the atoms or molecules, and a diminution in the effective volume of the atoms or molecules themselves. The first effect is almost the only effect in the compression of gases, and is by far the larger part of the effect in the compression of liquids up to a few thousand atmospheres. At higher pressures, diminution of volume is predominantly brought about by compression of the atoms or molecules themselves. This compression, or distortion, of the atoms results in a gradual disintegration of the electronic orbits within the atom until eventually, according to theory, the atomic constitution is completely lost and the substance becomes an amorphous sea of atomic nuclei and electrons.

the range up to a few hundred thousand atmospheres, in which substances retain approximately their normal atomic structure, there is a wide range in the magnitude of the loss of volume, depending on the substance. The volume loss ranges from a few per cent for the most incompressible (diamond) to more than 50% for the alkali metals and the solidified noble gases.

Changes of state. Perhaps the next simplest effect is the discontinuous change of volume which accompanies a change of phase. The best known of these is the freezing from the amorphous phase (gaseous or liquid) to the solid crystalline phase. Except for a few abnormal substances, such as water, the abnormality of which disappears at sufficiently high pressure, the effect of pressure is in general to raise the freezing temperature. The best opinion at present seems to be that this raising of the freezing point proceeds indefinitely, so that any substance could probably be maintained in the crystalline condition, no matter how high the temperature, by a sufficiently high pressure.

In addition to the ordinary phase change between solid and liquid, a great many substances occur in more than one crystalline form, and pressure in general may exert a great effect in changing one crystalline modification into another. This phenomenon may be of profound geological significance. Perhaps the most dramatic example is the synthesis of diamond out of graphite by pressure in the 100,000-atm range. There are no general rules governing the change of crystalline form under pressure, except for the one thermodynamic necessity that the stable high-pressure phase must have a smaller volume than the phase that is stable at lower pressure.

Changes in electrical conductivity. Another property affected by pressure is electrical conductivity. The effect of pressure on metals is, in the majority of cases, to increase electrical conductivity, but there are many metals in which the conductivity is decreased. Changes in the conductivity of metals are of the order of tens of per cent for pressures of tens of thousands of atmospheres. For other substances, such as semiconductors, the effects are often much larger.

Increase in viscosity. The largest of all pressure effects is that on the viscosity of liquids. Viscosity universally increases under pressure, and for some liquids, the increase may be millionfold for a pressure increase of 10,000 atm. See HIGH-PRESSURE PHENOMENA; HIGH-PRESSURE PROCESSES. [P.W.B.]

Bibliography: P. W. Bridgman, Recent work in the field of high pressures, *Rev. Mod. Phys.*, 18(1): 1–93, 1946; P. W. Bridgman, *The Physics of High Pressure*, 1949; E. W. Comings, *High Pressure Technology*, 1956.

High-pressure processes

Changes in the chemical or physical state of matter subjected to high pressure. The earliest high-pressure chemical process of commercial importance was the Haber synthesis of ammonia from hydrogen and nitrogen developed in Germany prior to World

War I. The synthesis of diamonds from graphite developed in the early 1950s is a high-pressure physical process. Raising the pressure on a system may result in several kinds of change. It causes a gas or vapor to become a liquid, a liquid to become a solid, a solid to change from one molecular arrangement to another, and a gas to dissolve to a greater extent in a liquid or solid. These are physical changes. A chemical reaction under pressure may proceed in such a fashion that at equilibrium more of the product forms than at atmospheric pressure; it may also take place more rapidly under pressure; and it may proceed selectively, resulting in the formation of more of the desired product among multiple possible products.

Pressures higher than that of the atmosphere are commonly expressed in atmospheres as well as in other units. The international unit atmosphere is defined as the pressure exerted by a column of mercury at 0°C, 76 cm in length, in a gravitational field giving an acceleration of 980.665 cm/sec^2.

Physical processes. Increasing the pressure on a gas or vapor compresses it to a higher density and so to a smaller volume. If the pressure exceeds the vapor pressure, the vapor will condense to a liquid which occupies a still smaller volume. A vapor may be condensed at a higher temperature when it is under pressure; this permits the use of cooling water to remove the latent heat instead of more costly refrigeration.

Solids also change from a less dense phase to a more dense phase under the influence of increases in pressure. The density of diamond is about 1.6 times greater than that of graphite because of a change in the spatial arrangement of the carbon atoms. The temperatures and pressures used in the commercial synthesis of diamond range up to 3,000°K and 100,000 atm. A molten metal is required as a catalyst to permit the atomic rearrangement to take place at economical rates of conversion. Metals such as tantalum, chromium, and iron form a film between graphite and diamond.

The highest static pressure attained in the laboratory is about 400,000 atm. At still higher pressures, the electrons are stripped from the atomic nuclei and matter loses its identity as recognizable atoms. This situation apparently exists in the centers of the white dwarf stars where pressures of the order of 10^{14} atm prevail.

Chemical processes. In a manner similar to its effect during a physical change in which the volume of a system decreases, pressure also favors a chemical change where the volume of the products is less than the volume of the reactants. This is Le Chatelier's principle, which applies to systems in equilibrium. This general principle may be derived more precisely by thermodynamic reasoning, and thermodynamics is used to predict the effect of pressure on physical and chemical changes which lead to an equilibrium state.

Ammonia production. Ammonia is formed according to the reaction

$$N_2 + 3H_2 \rightleftharpoons 2NH_3$$

The ammonia content at equilibrium in a mixture which contains initially a ratio of 3 moles of hydrogen to 1 mole of nitrogen is shown in Fig. 1. At one atmosphere, only a fraction of 1% ammonia is formed. The ammonia content increases greatly when the pressure is raised. At 100 atm and 200°C there would be about 80% ammonia at equilibrium. However, a very long time is required to form ammonia under these conditions, and consequently commercial processes operate at higher temperatures and pressures and use a catalyst to obtain higher rates of reaction. Many combinations of pressure and temperature have been used. The largest number of plants now operate in the region of 300 atm and 450–500°C. A higher pressure process is carried out at about 1000 atm and 500–650°C.

A catalyst must be used or the reaction is too slow. Thus the process does not operate at 100 atm and 200°C in spite of the favorable conversion at equilibrium because no catalyst has been found to provide an adequate rate of reaction under these conditions. Iron is the catalyst used at higher temperatures and pressures. The physical form of the catalyst is very important. Iron shavings or lumps are not effective. A suitable catalyst has been made from a fused mixture containing 66% Fe_2O_3, 31% FeO, 1.0% K_2O, and 1.8% Al_2O_3. The last two components are promoters and significantly increase the activity of the catalyst, but they alone are not catalysts. The mixture must not be contaminated with other impurities or the catalyst may be poisoned. Sulfur, phosphorus, and arsenic poison the catalyst permanently, whereas gases such as oxygen, water vapor, and carbon oxides reduce the catalyst's activity only while they are in contact with it and are temporary poisons. Before the catalyst is used, the iron oxides are first reduced to iron by the hydrogen of the reacting mixture. The catalyst is activated by this treatment and must not be exposed to gases which may contain poisons. Its life may be from several months to several

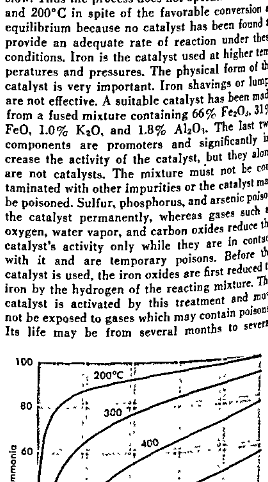


Fig. 1. Effect of temperature and pressure on formation of ammonia. (From E. W. Comings, *High Pressure Technology*, McGraw-Hill, 1956)

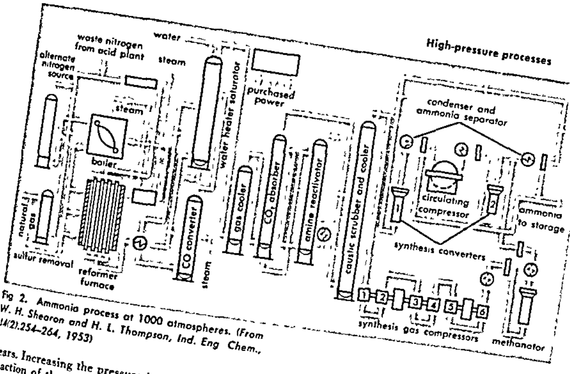


Fig 2. Ammonia process at 1000 atmospheres. (From W. H. Shearon and H. L. Thompson, Ind. Eng Chem., 14(2).254–264, 1953)

years. Increasing the pressure increases the rate of reaction of the mixture in contact with a suitable catalyst.

A flow sheet for an ammonia process operating at 1000 atm is shown in Fig. 2. Nitrogen is obtained from the air, and hydrogen comes from natural gas. The process involves preparation of the nitrogen-hydrogen mixture, purification to exclude catalyst poisons, compression to high pressure, and circulation of the synthesis gas through a closed loop which contains two synthesis converters charged with catalyst. Because of the incomplete conversion in one converter, the unreacted gas is passed on to the next converter after ammonia has been condensed from the partially reacted gas mixture.

The temperature of the reaction is carefully controlled. Higher temperatures shorten the life of the catalyst and reduce the degree of conversion at equilibrium, whereas lower temperatures reduce the rate of reaction. The reaction mixture entering the converter is relatively cool. It is heated to the reaction temperature inside the converter by heat exchange with the reacting gas and with the gas that has passed through the catalyst. The reaction is exothermic, and the converter is designed to remove some of the heat of reaction while the gas is passing through the catalyst.

The converters are large pressure vessels containing a heat exchanger and a catalyst basket. The converter for the 1000-atm process contains 16.5 ft^3 of catalyst and produces 250–400 lb of ammonia per hour in each cubic foot of catalyst. This is comparable to 144 ft^3 of catalyst in a converter for a 300-atm process, which produces 50 lb of ammonia per hour per cubic foot of catalyst. The high reaction temperature is confined to the catalyst basket inside the vessel, and the thick walls of the vessel are held at a lower temperature.

Methanol production Methanol is synthesized from hydrogen and carbon monoxide at 200 atm and 600°F in a similar manner. The catalyst contains aluminum oxide, zinc oxide, chromium oxide, and copper Higher alcohols are produced at pressures of 200–1000 atm and temperatures up to 1000°F with a similar catalyst to which potassium carbonate or chromate has been added.

Polyethylene production. Polyethylene has been produced at pressures up to about 2000 atm. This is probably the highest pressure yet used in the commercial synthesis of an organic chemical product. The ethylene is polymerized in a stainless steel tubular reactor at 375°F with small amounts of oxygen as a catalyst.

Phenol production. Phenol is formed from chlorobenzene mixed with 18% sodium hydroxide solution at a pressure of 330 atm. Pressure is employed in this instance to maintain the mixture in the liquid phase at a temperature high enough for the hydrolysis reaction to proceed at an acceptable rate. A tubular reactor installed in a furnace provides for heating the mixture to 680°F.

Other chemical reactions are carried out at elevated pressure. A reaction which has been observed to take place in the laboratory under high pressure may later be found to take place at lower pressure with a suitable catalyst.

Apparatus. The design of high-pressure apparatus must provide for the measurement of pressure, temperature, and the like; a vessel with sufficient strength, a closure to prevent leakage, and easy access to the interior; the compression of gases and liquids at high pressures; and the safety of people and equipment in case of ruptures or explosions. The dead-weight gage is the principal primary gage for calibrating other pressure-measuring instruments. It consists of a free piston balanced between the force of gravity acting on weights at one

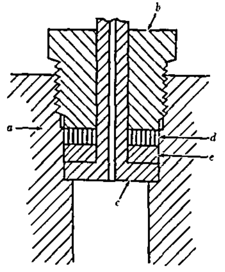


Fig 3. Bridgman unsupported area gasket. *(From E. W. Comings, High Pressure Technology, McGraw-Hill, 1956)*

end and the pressure of oil contained in the system at the other end.

P. W. Bridgman received the Nobel prize in 1946 for his pioneering work in the physics of high pressure. He was the originator of a self-sealing closure based on the unsupported area principle. This principle is illustrated in Fig. 3. The sealing gasket is initially under a low or moderate pressure. As pressure rises in the vessel, the pressure in the gasket is automatically maintained at a level higher than in the vessel. This ensures against leakage of the contents. Pressure, contained in the vessel *a*, acts upward on the steel disk *c*, which in turn acts on the gasket *e*. The gasket is supported by the steel ring *d*, which has a smaller area than *c* because of the unsupported area in the center. A balance of forces on the gasket requires that the forces acting downward balance those acting upward. The upward force is the product of the pressure in the vessel and the area of the lower face of *c*. This is balanced by an equal force downward, which is the product of the pressure in the gasket and the area of the gasket. Because the area of the gasket is smaller than the area of the lower face of *c*, the pressure in the gasket must always be higher than that in the vessel. *See* EQUILIBRIUM, CHEMICAL; EQUILIBRIUM, PHASE; HIGH-PRESSURE PHYSICS; UNIT PROCESSES. [E.W.CO.]

Bibliography: P. W. Bridgman, *The Physics of High Pressure*, reprint, 1949; E. W. Comings, *High Pressure Technology*, 1956; S. D. Hamann, *Physico-Chemical Effects of Pressure*, 1957.

Highway engineering

Before the design and construction of a new highway or highway improvement can be undertaken there must be general planning and consideration of financing. As part of general planning it is decided what the traffic needs of the area will be for a considerable period, generally 20 years, and what construction will meet those needs. To assess traffic needs the highway engineer collects and analyzes information about the physical features of existing facilities, the volume, distribution, and character of present traffic, and the changes to be expected in these factors (*see* TRAFFIC ENGINEERING). He must determine the most suitable location, layout, and capacity of the new routes and structures. Frequently, a preliminary line, or location, and several alternate routes are studied. The detailed design is normally begun only when the preferred location has been chosen.

In selecting the best route careful consideration is given to the traffic requirements, terrain to be traversed, value of land needed for the right-of-way, and estimated cost of construction for the various plans. The photogrammetric method, which makes use of aerial photographs, is used extensively to indicate the character of the terrain on large projects, where it is most economical. On small projects ground-mapping methods are preferred. *See* SURVEYING; TOPOGRAPHIC SURVEYING AND MAPPING.

Financing considerations determine whether the project can be carried out at one time or whether construction must be in stages, with each stage initiated as funds become available. In deciding the best method of financing the work, the engineer makes an analysis of whom it will benefit. Improved highways and streets benefit, in varying degrees, three groups: users, owners of adjacent property, and the general public.

Users of improved highways benefit from decreased cost of transportation, greater travel comfort, increased safety, and saving of time. They also obtain recreational and educational benefits. Owners of abutting or adjacent property may benefit from better access, increased property value, more effective police and fire protection, improved parking along the street, greater safety for pedestrian traffic, and the use of the street right-of-way for the location of public utilities, such as water lines and sewers.

Evaluation of the various benefits from highway construction is often a difficult problem and is an important part of highway engineering. Some benefits can be measured with accuracy, but the evaluation of others is more difficult. As a result numerous methods are used to finance construction, and much engineering work may be involved in selecting the best procedure.

Detailed design. Detailed design of a highway project includes preparation of the drawings or blueprints to be used for construction. These plans will show, among other things, the exact location, the dimensions of such elements as roadway width, the final profile for the road, the location and type of drainage facilities, and the quantities of work involved, including earthwork and surfacing.

Soil studies. In planning the grading operations the design engineer considers the type of material to be encountered in excavating or in cutting away the high points along the project and how the material removed can best be utilized for fill or for constructing embankments across low areas elsewhere on the project. For this the engineer must

g. 1. Large-scale earth-moving and construction of costly bridge structures were required for this section of a freeway north of Oakland, California.

analyze the gradation and physical properties of the soil, determine how the embankments can best be compacted, and calculate the volume of earthwork to be done. Electronic calculating procedures are now sometimes used for the last step. Electronic equipment has also speeded up many other highway engineering calculations.

In recent years powerful and highly mobile earth-moving machines have been developed to permit rapid and economical operations. For example, now in use in the United States is a self-propelled earth-mover weighing 125 tons and capable of hauling 100 yd^3 of earth. This unit is powered by a 600-hp diesel electric motor.

Surfacing. Selection of the type and thickness of roadway surfacing to be constructed is an important part of design. The type chosen depends upon the maximum loads to be accommodated, the frequency of these loads, and other factors. For some routes the traffic volume may be so low that no surfacing is economically justified and natural soil serves as the finished roadway. As traffic increases, a surfacing of sandy clay, crushed slag, crushed stone, caliche, crushed oyster shells, or a combination of these materials may be applied. If gravel is used, it usually contains sufficient clay and fine material to help stabilize the surfacing. Gravel surfaces may be further stabilized by application of calcium chloride, which also aids in controlling dust. Another surfacing is composed of portland cement and water mixed into the upper few inches of the subgrade and compacted with rollers. This procedure forms a soil-cement base that can be surfaced with bituminous materials. Roadways to carry large volumes of heavy vehicles must be carefully designed and made of considerable thickness. See PAVEMENT.

Drainage structures. Much of highway engineering is devoted to the planning and construction of facilities to drain the highway or street and to carry streams across the highway right-of-way.

Removal of surface water from the road or street is known as surface drainage. It is accomplished by constructing the road so that it has a crown and by sloping the shoulders and adjacent areas so as to control the flow of water either toward existing natural drainage, such as open ditches, or into a storm drainage system of catchbasins and underground pipes. If a storm drainage system is used, as it would be with city streets, the design engineer must give consideration to the total area draining onto the street, the maximum rate of runoff expected, the length of the design storm, the amount of ponding allowable at each catchbasin, and the proposed spacing of the catchbasins along the street From this information the desired capacity of the individual catchbasin and the size of the underground piping network are calculated.

In designing facilities to carry streams under the highway the engineer must determine the area to be drained, the maximum probable precipitation over the drainage basin, the highest expected runoff rate, and then, using this information, must calculate the required capacity of the drainage structure. Generally designs are made adequate to accommodate not only the largest flow ever recorded for that location but the greatest discharge that might be expected under the most adverse conditions for a given number of years.

Factors considered in calculating the expected flow through a culvert opening include the size, length, and shape of the opening, the roughness of the walls, the shape of the entrance and downstream end of the conduit, the maximum allowable height of water at the entrance, and the water level at the outlet.

There is a trend to use designs that permit drainage structures to be assembled from standard sections manufactured at a central yard. Such measures permit better control of the work, quicker construction, and less field work. For example, with precast concrete pipe sections it is often possible to avoid building small box culverts in the field. Circular culverts of large diameter are now also constructed in this way. However, culverts built of corrugated metal are specified for many projects for economy and to avoid placing small volumes of concrete at numerous locations.

Numerous small bridges are being designed to permit precast beams or girders to be placed side by side across the bridge opening to form the support for the roadway. These members are frequently of prestressed concrete. When precast members are used, the need for falsework to construct the bridge deck is eliminated, an especially beneficial move if the bridge is being built over a railroad or busy street. See PRECAST CONCRETE; PRESTRESSED CONCRETE.

Construction operations. Although much engineering and planning must be done preliminary to it, the actual construction is normally the costliest part of making highway and street improvements.

Staking out. With the award of a construction contract following the preparation of the d...

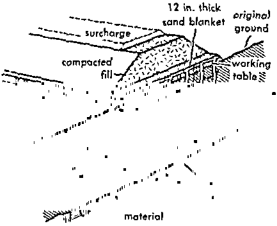


Fig. 2. Cross section of a vertical sand drain installation. Drains are 13–28 in. in diameter and spaced 9 ft apart. Dotted lines are staggered sand drains. (*From L. I. Hewes and C. H. Oglesby, Highway Engineering, Wiley, 1954*)

plans and specifications, engineers go onto the site and lay out the project. As part of this staking out, limitations of the earthwork are shown, location of drainage structures indicated, and profiles established.

Compaction. Heavy rollers are used to compact the soil or subgrade below the roadway in order to eliminate later settlement Pneumatic-tired rollers and *sheepsfoot rollers* (steel cylinders equipped with numerous short steel teeth or feet) are often employed for this operation. Vibratory rollers have been developed and used on some projects in recent years One type vibrates up to 3100 times/min, compacting the underlying material to an appreciable depth.

Vertical sand drains are sometimes employed to help stabilize fills or embankments constructed across wet and unstable ground (Fig. 2). Holes are drilled into the existing material and filled with sand. A horizontal layer of pervious sand or gravel is laid over the network of drains. As the fill is added over the pervious layer and the drains, the undesired water is forced upward through the soil. [illegible] roadway.

To obtain quicker compaction of the unstable soil, a surcharge may be placed above the compacted fill. The surcharge material, usually intended for building the compacted fill elsewhere on the project, is removed once the unstable soil has been sufficiently compacted.

Maintenance and operation. Highway maintenance consists of the repair and upkeep of surfacing and shoulders, bridges and drainage facilities, signs, traffic control devices, guard rails, traffic striping on the pavement, retaining walls, and side slopes. Additional operations include ice control and snow removal. Because it is valuable to know why some highway designs give better performance and prove less costly to maintain than others, engineers supervising maintenance can offer valuable guidance to design engineers. Consequently, maintenance and operation are important parts of highway engineering. [A.N.C.]

Hill and mountain terrain

Land surfaces characterized by roughness and strong relief. The distinction between hills and mountains is usually one of relative size or height, but the terms are loosely and inconsistently used Because of the prevalence of steep slopes, hill and mountain lands offer many difficulties to human occupancy. Cultivable land is scarce and patchy in occurrence, and transportation routes are often difficult to construct and maintain. The higher mountain ranges, by forcing moist air to rise in passing over them, induce large-scale condensation and precipitation. They also set up major disturbances in the broad pattern of atmospheric circulation and thus affect climates over extensive areas. Within the rough lands, large differences in elevation and in exposure to sunlight and to wind produce a complex pattern of local climatic contrasts.

Development of rough terrain. Mountain areas generally reflect disturbed structures in a portion of the earth's crust that has been subjected to strong folding, buckling, doming, or other disordering of rock structures. Sometimes the most intense crustal deformation is known to have occurred in the distant geologic past. However the existence of elevated ranges, on which the destructive work of erosion has only begun, is a clear indication that these are areas in which crustal deformation has continued to recent times, though perhaps less violently than before. Further evidence of crustal instability lies in the fact that the cordilleran belts are currently foci of earthquakes and volcanic activity. Hill lands, with their lesser relief, indicate only lesser uplift, not a fundamentally different course of development. Some hill and low-mountain land, however, is carved by streams cutting valleys into uplifted but little-distorted rock structures.

The elevated portions of the crust may have the form of broad, warped swells, somewhat smaller domes or arched folds, upthrust broken blocks, or folded and broken masses of extreme complexity. Some definitely limited areas owe their altitude to eruptive activity that has poured out thick sheets of lava or ejected immense quantities of volcanic ash.

Although crustal uplift is necessary in order to give mountain and hill lands their elevation, most details of surface character of these lands are erosional in origin, having been carved out of the uplifted masses by streams and glaciers. The principal exceptions are fresh volcanic cones and the occasional scarps and swells that have been produced by unusually rapid deformation of the crust in geologically recent time.

Distribution pattern of rough land. Although complex, rough-land distributions are not wholly without system. Hill and mountain terrain occupies

Fig. 1. The head of a glaciated mountain valley, showing a cirque and a steplike valley bottom, with lakes and waterfalls. (Photograph by Hileman from Glacier National Park)

about 36% of the earth's land area. The greater portion of that amount is concentrated in the great cordilleran belts that surround the Pacific Ocean, the Indian Ocean, and the Mediterranean Sea (see TERRAIN AREAS, WORLD-WIDE). Additional rough terrain, generally low mountains and hills, occurs outside the cordilleran systems in eastern North and South America, northwestern Europe, Africa, and western Australia. Eurasia is the roughest continental land, more than half of its total area and most of its eastern portion being hilly or mountainous. Africa and Australia lack true cordilleran belts; their rough lands are scattered about in patches and interrupted bands that rarely show marked complexity of geologic structure.

Predominant surface character. The features of hill and mountain lands are chiefly valleys and divides produced by sculpturing agents, especially running water and glacier ice. Local peculiarities in the form and pattern of these features reflect the arrangement and character of the rock materials within the upraised crustal mass that is being dissected.

Stream-eroded patterns. Stream-eroded features impart much of the character to most hill and mountain landscapes. In common with stream-sculptured surfaces of lesser relief, the principal differences from place to place are in the size, cross-section form, spacing and pattern of occurrence of the stream valleys, and the divides between them. These differences reflect variations in the original form and structure of the uplifted mass and in the stage to which erosion has proceeded.

In consequence of the uplift, streams have a large range of elevation through which they can cut, and as a result usually possess steep gradients early in their course of development. At this stage they are very swift, have great erosive power, and as a rule are marked by many rapids and falls. Because of the rapid downcutting, the valleys are characteristically canyonlike with steep walls and with floors no wider than the stream itself. At this same youthful stage of erosional development, divides are likely to be high and continuous, and broad ridge crests not uncommon. In hill lands such ridges often provide easier routes of travel than do the valleys, whereas in high mountain country the combination of gorgelike valleys and continuous high divides makes crossing unusually difficult. Thus in the Ozark Hills of Missouri, most of the highways and railroads follow the broad ridge crests. In the Himalaya and the central Andes, and to a lesser degree in the Rocky Mountains of Colorado, the narrow canyons are very difficult of passage, and the divides are so continuously high as to provide no easy pass routes across the ranges.

As erosional development continues, the major streams achieve gentle gradients and their valleys begin to widen. Divides become narrower and deep notches develop at valley heads, with well-defined peaks remaining between them. At this stage, which is fairly well represented by much of the Alps and by the Cascade Mountains of Washington, the principal valleys and the relatively low passes at their

heads furnish routes of no great difficulty across the mountain belts. If other conditions are favorable, the wide valleys may afford significant amounts of cultivable or meadow land, as is true in the Alps. Still further erosion continues the widening of valleys and reduction of divides until the landscape approaches the state of an erosional plain upon which stand only small ranges and groups of mountains or hills. The mountains of New England and many of the mountains and hill groups of the Sahara and of western Africa represent late stages in the erosional sequence.

Glaciated rough terrain. Glacial features of mountain and hill lands may be produced either by the work of local glaciers in the mountain valleys or by overriding glaciers of continental size. Continental glaciation has the general effect of clearing away crags, smoothing summits and spurs, and depositing debris in the valleys. The resulting terrain is less angular than is usual for stream-eroded hills, and the characteristic glacial trademark is the numerous lakes, most of them debris-dammed, a few occupying shallow, eroded basins. Examples of rough lands overriden by ice sheets are the mountains of New England, the Adirondacks, the hills of western New York, the Laurentian Upland of eastern Canada, and the Scandinavian Peninsula.

In contrast, mountains that have been affected by local glaciers are made rougher by the glacial action The long tongues of ice that move slowly down the valleys are excellent transporters of debris, and like their more extensive counterparts they are able to erode actively on shattered or weathered rock material. Valleys formerly occupied by glaciers are characteristically steep-walled and relatively free of projecting spurs and crags, with numerous broad cliffs, knobs, and shoulders of scoured bedrock. At their heads they generally end in cliff-walled amphitheaters called cirques.

The valley bottoms commonly exhibit steplike profiles, with stretches of gentle gradient alternating with abrupt rock-faced risers. Especially in the

that run lengthwise along the valley sides or swing in arcs across the valley floor. Lakes are strung

by the ice or that attendant upon the exposure of the rock surface by continual removal of the products of weathering, glaciated mountains are likely to be unusually rugged and spectacular. This is true not only of such great systems as the Himalaya, the Alps, the Alaskan Range, and the high Andes, but also of such lesser ranges as those of Labrador, the English Lake District, and the Scottish Highlands.

Most of the higher mountains of the world still bear valley glaciers, though these are not as large or as widespread as formerly. Dryness and long warm summers limit glaciers in the United States to a few large groups on the higher peaks of the

Fig. 2. The upper reaches of Susitna Glacier, Alaska, showing cirques, snowfields, and debris-laden glacier tongues. (*Photograph by Bradford Washburn*)

F … ograph by John L. Rich from … Appalachians of K … Geographical Review)

Pacific Northwest and numerous small ones in the northern Rockies.

Effects of geologic structure. Form and extent of the elevated areas, pattern of erosional valleys and divides, and, to some extent, sculptured details of slope and crest reflect geologic structure.

Some areas, such as the Ozarks, the western Appalachians, and the coast ranges of Oregon, are simply upwarped plains of homogeneous rocks that have been carved by irregularly branching streams into extensive groups of hills or mountains. Others, like the Black Hills or the ranges of the Wyoming Rockies, are domes or arched folds, deeply eroded to reveal ancient granitic rocks in their cores and upturned younger stratified rocks around their edges. The Sierra Nevada of California is a massive block of the crust that has been uplifted and tilted toward the west so that it now displays a high abrupt eastern face and a long canyon grooved western slope. The central belt of the Appalachians displays long parallel ridges and valleys that have been hewn by erosion out of a very old structure of parallel wrinkles in the crust. The upturned edges of resistant strata form the ridges; the weaker rocks between have been etched out to form the valleys. The Alps and Himalaya are eroded from folded and broken structures of incredible complexity involving almost all varieties of rock materials. Local variations in form and pattern reflect the differences in underlying material.

Most volcanic mountains, like the Cascades of the northwestern United States or the western Andes of Peru and Bolivia, are actually erosional mountains sculptured from thick accumulations of lava and ash. In these areas of volcanic activity, however, individual eruptive vents give rise to volcanic cones that range from small cinder heaps to tremendous isolated mountains. The greater cones, such as Fuji, Ararat, Mauna Loa, or Shasta, are among the most magnificent features of the earth's surface.

[E.H.HA.]

Hip

The region of the junction of the thigh and trunk. The bony framework consists of the large, flaring hip bone, or ilium, which contains the hip socket, or acetabulum. This receives the head of the thigh bone, or femur. Large muscle groups surround the joint and pass from trunk to thigh to impart motion to each. In addition, the region is rich in blood vessels, nerves, and lymphatics, many of which are of importance in the supply of the lower limb.

The term hip is used sometimes to designate the hip joint itself, and also to denote the lateral portions of the upper thigh. See SKELETAL SYSTEM.

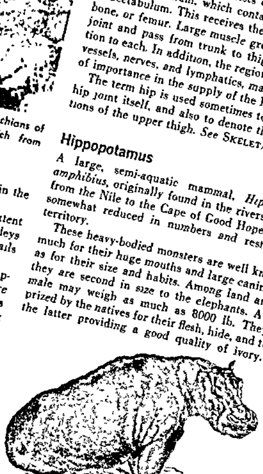
[E.G.ST.]

Hippopotamus

A large, semi-aquatic mammal, *Hippopotamus amphibius*, originally found in the rivers of Africa from the Nile to the Cape of Good Hope, but now somewhat reduced in numbers and restricted in territory.

These heavy-bodied monsters are well known, as much for their huge mouths and large canine teeth as for their size and habits. Among land animals, they are second in size to the elephants. A large male may weigh as much as 8000 lb. They are prized by the natives for their flesh, hide, and tusks, the latter providing a good quality of ivory. Although they may damage crops, the favorite food of hippopotamuses is rooted aquatic plants which abound in the quiet tropical waters where they live. See ARTIODACTYLA.

[J.D.B.]



The hippopotamus, *Hippopotamus amphibius*; length to 17 ft. (From P. M. Duncan, ed., Cassell's Natural History, Cassell)

Hirudinea

A class of the annelid worms commonly known as the leeches. These organisms are parasitic or predatory and have terminal suckers for attachment and locomotion. Most inhabit inland waters, but some are marine and a few live on land in damp places. The majority feed by sucking the blood of other animals, including man.

Morphology. Leeches differ from other annelids in having the number of segments in the body fixed at 34, chaetae or bristles lacking, and the coelomic space between the gut and the body wall filled with packing tissue (*see* ANNELIDA). In a typical leech, the first six segments of the body are modified to form a head, bearing eyes and a sucker, and the last seven segments are incorporated into a posterior sucker. Each segment is divided externally into 2–16 rings or annuli. The number per segment is constant for a particular species in the midbody region, but decreases toward the extremities.

The mouth of a leech opens within the anterior sucker, and there are two main methods of piercing the skin of the host to obtain blood. The rhynchobdellid leeches have the anterior part of the gut modified to form an eversible proboscis (*see* RHYNCHOBDELLAE). The anterior sucker is planted firmly in position on the skin of the host, and the proboscis is forced out of the leech's mouth and into the host tissue. The arhynchobdellid leeches lack the proboscis, and in its place have three jaws, each shaped like half a circular saw, placed just inside the mouth (*see* ARHYNCHOBDELLAE). When the anterior sucker is placed in position on the skin of the host, the three jaws are pushed forward and rocked in such a way as to cause a Y-shaped incision. Salivary glands are present in both orders of the leeches, and their secretion helps to prevent the clotting of blood. The jawed leeches are able to penetrate the skin of man and other mammals, but a proboscis is less efficient, and rhynchobdellid leeches usually have to seek out softer tissues such as the lining of the nostril of a bird, or the gills of a fish. Some arhynchobdellid leeches have given up bloodsucking and feed instead on such animals as earthworms and insects. In these leeches, the jaws are often reduced to muscular ridges used merely to grasp the prey, normally swallowed whole.

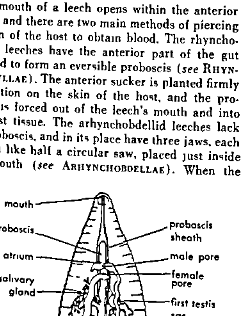


Diagram illustrating the general structure of a leech. The male reproductive system is shown on the right, the female on the left. (From K. H. Mann, A key to the British freshwater leeches, Freshwater Biological Assoc. Sci. Publ., 14:3–21, 1954)

The bloodsucking leeches have a large region of the gut, the crop, modified for the storage of blood. It consists of a central chamber with a number of paired diverticula, or side branches. When a meal of blood is taken, the crop becomes considerably distended, so that some leeches can store up to ten times their own weight in blood. The process of digestion is very slow, and a meal may last a leech for 9 months. The carnivorous forms have lost most or all of their gut diverticula so that they resemble earthworms in possessing a straight, tubular gut.

Leeches are hermaphroditic, having a single pair of ovaries and several pairs of testes. The latter are arranged on either side of the gut, and the sperms pass forward in paired ducts to a single midventral opening placed about one-third of the distance from the anterior to the posterior sucker. In many leeches, there is an eversible penis, and sperms are transferred directly to the female pore of another leech. However, in others the penis is lacking, and the sperms are made up into packets, or spermatophores, which are attached to the body surface of another leech. From here the sperms migrate through the tissues to fertilize the eggs. The female pore is in the midventral line, a short distance behind the male pore, and at the time of egg laying a thickened, glandular region of the body wall, the clitellum, secretes a barrel-shaped cocoon around the body, over the genital pores. The fertilized eggs are passed into this cocoon, after which the leech works the cocoon over its head and then closes the ends. Some leeches loosely deposit the cocoons in a damp place, others cement them to stones or vegetation under water, while still others place them under their bodies and brood over them. In the last case the young, on hatching, attach themselves to the under surface of the parent and are carried about for some time. The method of reproduction is very like that of the earthworm (*see* OLIGOCHAETA), and leeches may be regarded as earthworms which have become modified for the parasitic mode of life. *Acanthobdella* is regarded as a transitional form. It has only 30 segments in the body, of which 5 bear bristles, or chaetae. There is no anterior sucker, and the coelomic space is not entirely filled by packing tissue.

Economic importance. The importance of leeches as a means of making incisions for the

letting of blood or the relief of inflammation is declining, and in civilized countries the bloodsucking parasites of mammals are declining, owing to lack of opportunity for contact with the hosts. In other countries they are still serious pests. *Limnatis nilotica* of the Middle East enters the mouth or nostrils of animals or humans drinking at streams, and causes bleeding, nausea, and vomiting, often followed by death from anemia or suffocation. The leech may sometimes be removed, with difficulty, by the use of irritants such as salt water, chloroform water, or vinegar. *Hirudo* is widespread in Europe and Asia. Its attacks on man and domestic animals are usually not serious, since they usually occur on the body or the limbs, but occasionally serious damage is caused by attacks on delicate tissues such as the eyes. In South America, the attacks of *Haemadipsa chiliani* on horses and cattle can be fatal. *Theromyzon*, which enters the throat and nostrils of water birds, may cause serious losses of domestic ducks and geese when crowded conditions lead to an increase of the population of the leech. Similarly, the fish parasite *Piscicola* may become a serious pest in hatcheries or heavily stocked fishponds.

On the other hand, many of the leeches which are carnivorous in habit form an important contribution to the diet of fresh-water fishes. Records from the stomach contents of fish do not often show this, owing to the rapid and complete digestion of the leeches. [K.H.M.]

Bibliography: W. A. Harding and J. P. Moore, *Hirudinea*, 1927; K. Herter, *Hirudinea*, in P. Schulze (ed.), *Biologie der Tiere Deutschlands*, 35(12b):1–158, 1932; K. H. Mann, The ecology of the British freshwater leeches, *J. Animal Ecol.*, 24(1):98–119, 1955

Histamine

A derivative of the amino acid histidine which is widely distributed in tissues. The effects produced by its release are manifold; almost every type of tissue or organ shows some response to it. Histamine is a powerful capillary dilator and may produce changes marked by loss of blood fluid and plasma protein from the vessels to adjacent tissue spaces. The so-called triple response found in many types of local reaction is thought to be at least partially due to the effects of released histamine or a histaminelike substance.

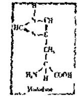
Histamine also causes either constriction or dilatation of the arteries and of particular portions of the circulatory system, depending on the animal species and other factors. See CARDIOVASCULAR SYSTEM.

This substance may also cause smooth muscle contraction, particularly of the uterus, the bronchioles, and the gallbladder. Many secreting glands are stimulated by histamine; this is the basis for the use of the substance in diagnostic tests of gastric secretion.

Histamine shock is a state of peripheral vascular dilatation that causes a rapid fall in blood pressure and other deleterious effects. *The similarity to the shock produced by trauma and certain acute allergic reactions has led to the belief that all are due to an excessive or sudden release of histamine; this theory is still controversial.*

Antihistamine compounds have been developed to combat the general and specific effect of histamine release. These antihistamines are largely symptomatic in action and do not remove the inciting cause of the histamine release. See ANTIHISTAMINES. [E.G.ST.]

Histidine

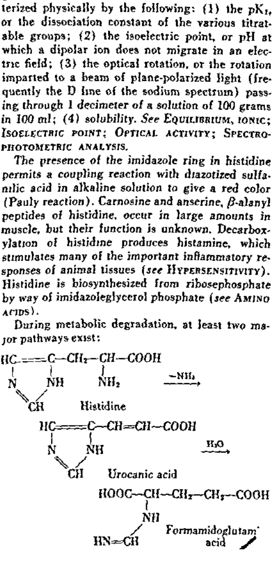


Physical constants of the L isomer at 25 C
pK_1 (COOH) 1.82 pK_2 (imidazole) 6.00, pK_3 (NH_3^+) 9.17
Isoelectric point 7.59
Optical rotation $[\alpha]_D(H_2O)$ −38.5, $[\alpha]_D$(5 N HCl) +11.8
Solubility (g 100 ml H_2O) 4.29

An amino acid considered essential for normal growth of animals. The amino acids are characterized physically by the following: (1) the pK_1, or the dissociation constant of the various titratable groups; (2) the isoelectric point, or pH at which a dipolar ion does not migrate in an electric field; (3) the optical rotation, or the rotation imparted to a beam of plane-polarized light (frequently the D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml; (4) solubility. See EQUILIBRIUM, IONIC; ISOELECTRIC POINT; OPTICAL ACTIVITY; SPECTROPHOTOMETRIC ANALYSIS.

The presence of the imidazole ring in histidine permits a coupling reaction with diazotized sulfanilic acid in alkaline solution to give a red color (Pauly reaction). Carnosine and anserine, β-alanyl peptides of histidine, occur in large amounts in muscle, but their function is unknown. Decarboxylation of histidine produces histamine, which stimulates many of the important inflammatory responses of animal tissues (*see* HYPERSENSITIVITY). Histidine is biosynthesized from ribosephosphate by way of imidazoleglycerol phosphate (*see* AMINO ACIDS).

During metabolic degradation, at least two major pathways exist:



1. Deamination to urocanic acid, followed by hydrolysis to α-formamido-L-glutamic acid. This compound is further degraded to glutamic acid, formate, and ammonia.

2. Decarboxylation to histamine, which is then oxidized, first to imidazoleacetaldehyde, and then to imidazoleacetic acid. This compound is oxidized to formylaspartic acid in bacteria; deformylation then yields aspartic acid.

[E.A.AD.]

Histogenesis

The developmental processes by which the definite cells and tissues comprising the body of an organism arise from embryonic cells. Among animals, the ectoderm, endoderm, and mesoderm, also known as the primary germ layers, provide the stem cells which gradually transform into distinctive kinds of cells and tissues. In the higher plants, meristematic cells, which occur wherever extensive growth takes place, provide the basis for tissue formation. Parenchyma, a simple permanent tissue, may dedifferentiate and become meristematic. *See* EMBRYOLOGY, HISTOLOGY. [C.B.C.]

Histology

The study of the structure and chemical composition of tissues of animals and plants as related to their function. The primary aim is to understand how tissues are organized at all structural levels including the molecular and macromolecular, the entire cell and intercellular substances, and the tissues and organs.

The four tissues of the animal body include cells *and* intercellular substances They are (1) epithelium, in which the cells are generally closely applied to each other and separated by very little *intercellular* substance; (2) connective tissue, in which the cells are usually separated by greater amounts of intercellular substance, which may indeed form the great bulk of the tissue; (3) muscu*lar tissue*, whose cells are primarily concerned with *contractility*; and (4) [illegible] gross anatomy.

The structures *studied in* histology extend from those which are just too small to be seen with the hand lens to those *which* are beyond the resolution of the electron microscope, for example those which measure *about* 20 A. For this purpose, a wide range of instruments has *been used*. These include x-ray [illegible] ele[illegible] fiel[illegible] ligh[illegible] phase contrast and interference microscope, and the dissecting *microscope*.

In earlier days, fresh material to be examined by the histologist was either *teased* or sliced free-hand, scraped, or smeared so that it would be thin enough to be viewed with a microscope by transmitted light. Later, microtomes were introduced which could cut sections as thin as 1 μ; now they are capable of cutting sections 70 A in thickness. The period of use of the microtome corresponds roughly with the use of fixatives, which were introduced to preserve and retain structure. The structural rearrangements, distortions, and losses caused by the use of fixatives and the use of irrelevantly esthetic stains eventually caused a reaction among histologists. This took the form of a return to the study of fresh material under strictly controlled conditions, the development of tissue culture and micromanipulation, microcinematography, intravital staining (of living cells), and supravital staining (of surviving cells). Another aspect of this reaction was the practical development of fixation by freezing and drying. In this method of preservation, specimens are frozen very rapidly by immersion in fluids cooled to −150°C or less, dried in vacuo at a temperature of −30°C or less, and later infiltrated in paraffin.

The major fields of histological studies are: (1) morphological descriptions; (2) developmental studies; (3) histo- and cytophysiology; (4) histo- and cytochemistry; and (5) fine (or submicroscopic) structure. Histo- and cytophysiology deal with correlations between morphological changes and functional activity. Histo- and cytochemistry deal with the chemical composition of morphological structures. The study of fine structure deals with the arrangements of structures which are below the resolution of the light microscope (about 0.2 μ). *See* CONNECTIVE TISSUE; EPITHELIUM; NERVOUS SYSTEM. *See also* ANGSTROM. [I.G.]

Bibliography: G. H. Bourne, *An Introduction to Functional Histology*, 1953; J. L. Bremer and H. L. Weatherford, *A Textbook of Histology*, 6th ed., 1944; R. O. Greep, *Histology*, 1954; A. W. Ham, *Histology*, 3d ed., 1957; A. A. Maximow and W. Bloom, *A Textbook of Histology*, 7th ed., 1957; J. F. Nonidez and W. F. Windle, *Textbook of Histology*, 2d ed., 1953.

Histoplasmosis

An infectious fungus disease primarily of man, although it has also been observed in animals. It is caused by a fungus, *Histoplasma capsulatum*, which has dimorphic properties. The infectious *form* is a mold whose spores are inhaled. As the organism invades tissues, the mold form is converted to a yeast form. This organism has been recovered *from* soil and certain plants. It is endemic to the valleys of the Mississippi and Ohio Rivers. *See* ASCOMYCETES; YEAST.

Histoplasmosis is primarily a disease of the lungs, although the disseminated form *may occur*, usually in young children and in adults *in their* fifth or sixth decades. Approximately 50% of those individuals with the pulmonary form are asymptomatic. Diagnosis is made by recovery of the organism from pathological tissue. Antigens are avail-

able for skin testing and serological studies. There is no satisfactory treatment for this disease. See ANTIGEN; MYCOLOGY, MEDICAL. [L.D.H.]

Historadiography

The technique for taking x-ray pictures of cells, tissues, or sometimes the whole animal or plant, if it is a small one. Soft x-rays, those with low penetrating power and relatively long wavelengths, are required for this type of picture. The best pictures are obtained when the tissues contain deposits of metallic elements which have a high absorption capacity for x-rays. See X-RAY(S), PHYSICAL NATURE OF.

In applying the technique to tissues, a relatively thin section is placed against an x-ray film and irradiated with a beam of x-rays. When the film is developed, a picture of the object or section of tissue shows on the film. Another method attempts to focus the x-rays after they pass through the specimen. In this arrangement, a lens is placed between the specimen and the film. X-rays are very difficult to focus. The lenses must be the reflecting type of mirror surfaces. A simple two-mirror system has two cylindrically curved surfaces set at right angles to each other. The glancing angle must be small to get reflection of the x-rays. The lenses also have certain aberrations that limit the resolution to about 0.2 μ, which is no better than the best light microscope images. However, x-rays with their short wavelengths are potentially capable of much higher resolution if they could be focused without much aberration. The advantage of the x-ray picture is that it may resolve detail not visible with light in calcified bone tissues or in tissues that have been injected with an x-ray–absorbing material.

An interesting use of historadiography in biology is the determination of the relative mass of cellular structures. The tissue is quick frozen and dried at low temperatures. Then sections are prepared and photographed by the method described above. A relation exists between the mass of the various parts of the specimen and the contrast of the film, that is, the number of silver grains or the amount of silver deposited. Quantitative determinations can be made by scanning the photograph with a densitometer. See NUCLEAR RADIATION (BIOLOGY). [J.H.T.]

Bibliography: G. Oster and A. W. Pollister (eds.), *Physical Techniques in Biological Research*, vol. 3, 1956.

Hoarding behavior

The carrying of food to the home nest for storage, in quantities exceeding daily need. It is a homeostatic behavior pattern with high survival value, assuring the hoarder a stable food supply independent of fluctuations in its territory. While many species store food, only hoarding in rats has been systematically studied. See HOMEOSTASIS.

The basic response of hoarding, retrieving a food pellet and taking it directly to the nest, is unlearned, for it appears in the rat's repertoire without previous experience with solid food.

Since rats collect many things besides food and since the pattern of response is the same for carrying any material to the nest, the distinctive features of hoarding are the kind of objects selected and the motivation underlying the animal's choice.

Motivation to hoard. Rats that have never suffered deprivation generally gather only a day's supply of food but, given the opportunity, they collect nuts, candy, and foil-wrapped pellets. Food-deprived rats hoard pellets; vitamin-B–deficient rats retrieve standard pellets in preference to pellets lacking that vitamin. If starved rats are returned to an adequate diet, in time they cease hoarding.

While food deprivation brings about hoarding, hunger is neither the necessary nor sufficient condition of the behavior. Periodic fasting has a cumulative rather than an immediate effect, and hoarding does not disappear until after several days of free feeding.

Exposure to cold induces active pellet storing, as an inverse function of environmental temperature. This is not entirely food hoarding, however. When more suitable insulating material is not available, rats will use food pellets to conserve body heat, as well as for food stores (see REPRODUCTIVE BEHAVIOR).

The amount of hoarding increases with the severity of the fasting schedule, the age of the animal, its past or present subjection to feeding frustration and psychological stress, and its familiarity with the available food and with the environment. Removing or adding to the rat's hoard has no effect upon the number of pellets that are carried.

Physiological and neural mechanisms. Distinct differences exist between strains of rats, in the onset and amount of hoarding during deprivation and satiation. When comparisons are made across strains, hoarding is positively correlated with activity level, fertility, and aggression. Thus, hoarding may depend on genetically determined structures and functions, perhaps hormonal, neural, or both. The relevant biochemical states and neural structures are still unknown. There is some evidence that lesions of the medial cortex reduce hoarding in accordance with the mass of the lesion.

Comparative studies suggest that different mechanisms may underlie food storing in different species. See INSTINCTIVE BEHAVIOR. [C.HO.]

Hodgkin's disease

A progressive, ultimately fatal disease of lymphoid tissue. It is characterized by variable enlargement of the lymph nodes and, frequently, of the spleen (splenomegaly). The age group of 20–40 is most often affected; the disease occurs two to three times as often in men as in women.

Histologically and clinically, three basic varieties are seen: the Hodgkin's paragranuloma, granuloma, and sarcoma. The microscopic feature common to all is the presence of the pathognomon-

Reed-Sternberg cell, named for investigators who described it.

Although Hodgkin's disease is generally considered to be a malignant neoplasm, there is some evidence for its origin from an infection.

Abnormal development and proliferation of the lymphoid cells in various parts of the body is the cause of the enlargement of lymphoid tissue and organs. Frequently there is invasion of the liver, bones, lungs, and other organs. Anemia, weight loss, and a characteristic type of relapsing fever (Pel-Ebstein fever) are common clinical findings.

Each of the three basic varieties has its own clinical and microscopic features but a shift to more rapidly progressive forms is often seen; the sarcoma is the most malignant type.

Radiation th...

yet available. See NUCLEAR RADIATION (BIOLOGY).

Hodgkin's disease is often grouped with other malignant lymphomas and certain types of leukemia since a common cell, the lymphocyte, is involved and the clinical courses are somewhat similar. See ONCOLOGY. [E.G.ST.]

Hog

An even-toed mammal, *Sus scrofa*, of the family Suidae, found as a domesticated animal in many strains throughout the world. It occurs in the wild state in Eurasia and has been introduced elsewhere, including the United States. In the feral state it is known as the wild boar.

The domesticated h...

tan...

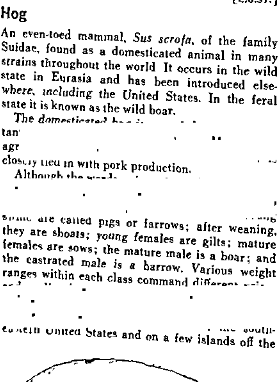
agr...

closely tied in with pork production.

Although the words ...

... are called pigs or farrows; after weaning, they are shoats; young females are gilts; mature females are sows; the mature male is a boar; and the castrated male is a barrow. Various weight ranges within each class command different ...

... southeastern United States and on a few islands off the



The Yorkshire hog, *Sus scrofa*. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

coast of California. This tough, lean, gray to blackish animal attains a length of over 5 ft, and may weigh 350 lb or more. Its tusks may exceed 1 ft in length. It is unquestionably the most dangerous wild animal in America. See ARTIODACTYLA; SWINE PRODUCTION. [J.D.B.]

Hog-nosed snake

Any of three species of the genus *Heterodon* characterized by a strong rostral plate on the tip of the snout which is keeled and turned upward. They are stout-bodied snakes, with keeled scales. When disturbed, they have the habit of flattening the head and forepart of the body in a threatening, cobralike manner, and hissing with the mouth open. They are completely harmless, although there is a widespread belief that they are deadly poisonous. They are known by many names including spreading adder, puff adder, and spreading viper. They feign death when their bluff fails them. These snakes are among the few predators that will eat toads, these being one of their staple items of diet. They also eat frogs, shrews, mice, and small snakes. Except for the far West, New England states, and the northern fringe, some form of hog-nosed snake is found throughout the United States. See SQUAMATA. [J.D.B.]

Hoisting machines

Mechanisms for raising and lowering materials with intermittent motion while holding them freely suspended. Hoisting machines are capable of picking up these loads at one location and depositing them at another within a limited area. In contrast, elevators move their loads only in fixed, vertical paths, and monorails operate over a fixed, normally horizontal path rather than over a limited area (see ELEVATING MACHINES; MONORAIL).

Sheaves and pulleys are used primarily as hoist ing tackle; winches and hoists are power units; derricks and cranes are structural elements which in various combinations with the first two groups are the principal components of hoisting machines.

Block and tackle. Sheaves and pulleys or blocks are a means of applying power through a rope, wire cable, or chain. Sheaves are grooved wheels in appropriate mountings which are used to change the direction or the point of application of a force. Pulleys are made up of one or more sheaves mounted in a frame, usually with an attaching hook, eye, or similar device at one or both ends.

Pulley systems are combinations of blocks. Mechanical advantage *MA* of a pulley system is the ratio of weight lifted W to input effort exerted E. Mechanical advantage, neglecting friction, for any of various arrangements can be determined readily because it equals the number of strands which support the load (Fig. 1). The ratio of distance D_E through which the effort moves in lifting the weight through distance D_W is inversely proportional to the mechanical advantage.

Sometimes used alone, sheaves and pulleys find their most usual application as the hoisting tackle of derricks and cranes.

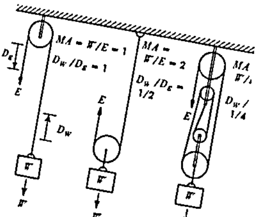


Fig 1. Mechanical advantage of pulley systems.

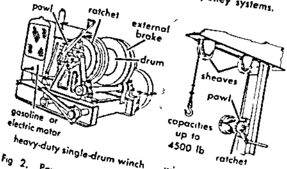


Fig 2. Powered and hand winches.

Winches and hoists. Winches are designed for stationary service; hoists are mounted so that they can be moved about, for example on wheel trolleys in connection with overhead crane operations.

Winches are constructed with a drum around which cordage is coiled for hoisting or hauling. The drum may be rotated manually or by power. A ratchet and pawl prevent the load from slipping; large winches are equipped with brakes, usually of the external band type. Industrial applications of winches include use as the power element for derricks and as the elevating mechanism with stackers (Fig. 2). Floor- and wall-mounted electric hoists are used for many hoisting and hauling jobs from fixed locations in industrial plants and warehouses. Heavy-duty types are standard equipment for powering ship's gear in cargo handling. Mounted on over-the-road carriers, they facilitate the moving of heavy bulky loads. They serve as the power units of power cranes and shovels. A car-puller is a special type, with the drum mounted vertically, used for spotting railroad cars in freight yards.

Hoists are designed to lift from a position directly above their loads, hence mobile mountings are necessary. An exception is a whip hoist which can lift from slightly offset positions. Hoists are classified by their source of power such as hand, electric, or pneumatic.

Hand hoists are chain actuated and have lever-ratchet, differential, spur- or worm-gear mechanisms (Fig. 3). Simplest, least expensive, and least efficient is the lever-operated ratchet type. Next is the differential type which depends for its action on the difference in the diameters of the two pulleys in the top block. The names screw- or worm-gear and spur-gear indicate the mechanical means used to operate the next two varieties. Spur-gear chain hand hoists have the highest first cost, but they are the most efficient, being rated as high as 85%. As useful as hand hoists are, they are restricted in their application to spotty operations where time is not important.

Electric hoists lift their loads either by cable or by chain (Fig. 4). The first types have a drum around which a wire cable is coiled and uncoiled for hoisting and lowering. Chain models have either a roller chain and sprocket or a link chain and wheel for hoisting and lowering.

Innumerable below-the-hook attachments such as slings, hooks, grabs, and highly specialized devices facilitate practically any handling requirement. Many of these devices are so designed that they pick up and release their loads automatically.

Pneumatic or air hoists are constructed with cylinders and pistons for reciprocating motion and as air motors for rotary motion. In both, compressed air is the actuating medium. Various arrangements admit air to and discharge it from the

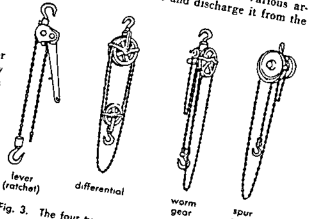


Fig. 3. The four types of chain hand hoists.

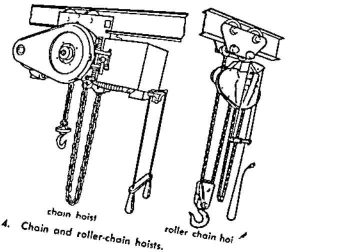


Fig. 4. Chain and roller-chain hoists.

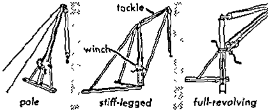


Fig. 5. Principal types of derricks. *(From D. O. Haynes, Materials Handling Equipment, Chilton, 1957)*

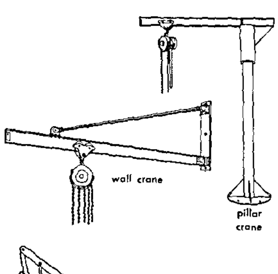


Fig. 6. Three types of jib cranes.

cylinders, which are mounted to operate vertically or horizontally. The smooth action of these hoisting machines and their sensitive response to control account for their wide use in handling fragile materials, such as molds in foundries. Freedom from sparking makes them useful in locations where the presence of explosive mixtures makes electrical [illegible]

[illegible] at its lower end and carrying load-supporting tackle at its outer end. In contrast, *jib cranes* always have horizontal booms, and slant-boom cranes operate without masts.

Derrick masts are supported by guy lines or *stiff legs;* some are arranged to rotate through 360° (Fig. 5). Winches, hand or powered, usually in conjunction with pulleys, do the lifting. Standard equipment on construction jobs, winches are frequently mounted on barges for lighterage and dredging operations.

Jib cranes, when carried on self-supporting masts, are called pillar cranes and those mounted on walls are wall bracket cranes. Cranes with jib-like booms are used in shops (Fig. 6). They may have their own running gear or be mounted on trucks. Mobile types for heavier service are called yard cranes or crane trucks. These may or may not be able to rotate their booms. More powerful machines belong to the power crane and shovel group (*see* BULK-HANDLING MACHINES).

Overhead-traveling and gantry cranes. Hoisting machines with a bridgelike structure which spans the area over which they operate are overhead-traveling or gantry cranes. In the former type, the bridge is carried by and moves along overhead trackage, whereas in the latter, the bridge is normally supported by fixed structures or arranged for running along tracks on the ground. There are variations in both types.

Basic arrangements of overhead-traveling cranes are top-running and underhung. In the former, the bridge's end trucks ride on top of the runway rails; in the latter, the end trucks carry the bridge sus-

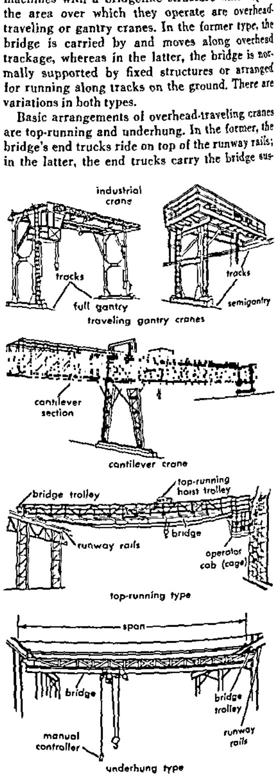


Fig. 7. Basic types of overhead-traveling cranes.

pended below the rails (Fig. 7). Types for relatively light duty can be made of elements used in the construction of overhead track (*see* MONORAIL). Bridges for heavier duty are constructed of one or more steel girders. Bridges and trolleys may be hand-propelled or electrically powered. In some instances, two different hoisting mechanisms are provided, one for light duty, the other for heavier service. Controls may be pendant-type push buttons operated from the ground or, in larger units, the operator may be located in a cab (cage). Radio communication between cab and yard aids in large operations. Operatorless cranes are now operable by means of electronic controls.

Full gantry cranes have both their supporting elements erected on the ground, while semigantry types have one leg running on the ground, the other on elevated trackage. A cantilever variation is widely used in marine terminal operations. Characteristically, the hoist trolleys run on a bridge which extends beyond the limits of the supporting legs. Other variations are rotating cranes and hammerhead types in which the hoisting mechanism rides on counterweighted runways.

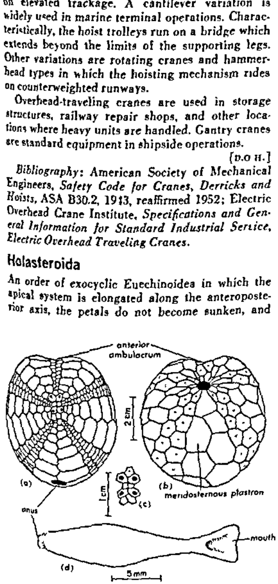
Overhead-traveling cranes are used in storage structures, railway repair shops, and other locations where heavy units are handled. Gantry cranes are standard equipment in shipside operations.

[D.O H.]

Bibliography: American Society of Mechanical Engineers, *Safety Code for Cranes, Derricks and Hoists*, ASA B30.2, 1943, reaffirmed 1952; Electric Overhead Crane Institute, *Specifications and General Information for Standard Industrial Service, Electric Overhead Traveling Cranes*.

Holasteroida

An order of exocyclic Euechinoidea in which the apical system is elongated along the anteroposterior axis, the petals do not become sunken, and



Diagnostic features of holosteroids. (*a*) Aboral aspect. (*b*) Adoral aspect. (*c*) Apical system. (*d*) *Echinosigra paradoxa*, a deep-sea species of Pourtalesiidae, adoral aspect. (After T. Mortensen)

teeth occur only in the juvenile stages (see illustration). There are seven families. The Collyritidae and Disasteridae were mainly small ovoid forms without fascioles or plastron. They burrowed in the sea floor in shallow or moderately deep water. They were exclusively Tethyan from the Jurassic and Early Cretaceous. In the five other families the posterior interambulacral plates behind the mouth became arranged in a single series, the meridosternous plastron, and fascioles developed (*see* SPATANGOIDA). The extinct Stenomasteridae and the Holasteridae are oval or heart-shaped forms, with fully developed pore pairs. They are mainly Cretaceous and Eocene and only a few holasterids survive. The other families have the pores unpaired aborally, and survive today as fragile deep-sea forms living in cold abyssal muds. The Pourtalesiidae have a bottle-shaped test. Urechinidae and Calymnidae have an ovoid test, the latter with a marginal fasciole which is lacking in the Urechinidae. *See* ECHINOIDEA; EUECHINOIDEA. [H.B.F.]

Holectypoida

An order of exocyclic Euechinoidea with keeled, flanged teeth, distinct genital plates, and with the ambulacra narrower than the interambulacra on the adoral side. Of the seven families, six are extinct. The Holectypidae, from the Jurassic and Cretaceous, were hemispherical. The small ovoid Conulidae were exclusively Cretaceous and had a flattened oral surface. Internal skeletal partitions had developed in the conical or globular Discoidiidae, and the Galeritidae were distinguished by their large ambulacral plates with small, widely separated pore pairs. Aboral petals developed in two families, namely, the Cretaceous and Eocene

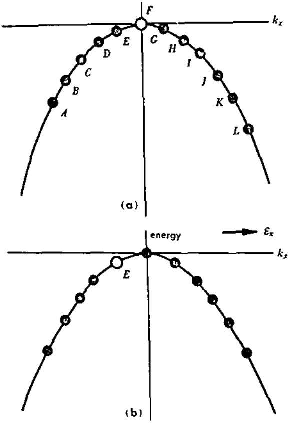
Africa. The last family, Echinoneidae, are small, round or elongate forms which arose in the Jurassic, reached a world-wide climax in the Cretaceous, and survive today in two genera, *Echinoneus* and *Micropetalon*. These live buried in mud or below coral slabs in the littoral zone of tropical seas. *See* ECHINOIDEA; EUECHINOIDEA. [H.B.F.]

Holes in solids

Vacant electron energy states near the top of an energy band in a solid are called holes. A full band cannot carry electric current; a band nearly full with only a few unoccupied states near its maximum energy can carry current, but the current

$$m^* = \hbar^2 \left(\frac{\partial^2 E}{\partial k^2}\right)^{-1}$$

where $\hbar$ is Planck's constant divided by 2π (see AND THEORY OF SOLIDS). Near a maximum of



(a) At time $t = 0$, all energy states A through L are filled except F at the top of the band. (b) An electric field $\mathcal{E}_x$ is applied in the $+x$ direction. The force on the electrons is in the $-k_x$ direction, and all electrons make transitions together in the $-k_x$ direction, moving the hole to E. (c) After a further interval, the electrons move farther along, and the hole is now at D. In this way, the hole has moved to decreasing k_x values, corresponding to higher positive velocities. (After C. Kittel, *Introduction to Solid State Physics*, 2d ed., Wiley, 1956)

band, the second derivative of the energy is negative, so the effective mass is negative. States for which the effective mass is negative are defined as hole states. Carriers in such states behave under the influence of an external electromagnetic field as though they carry positive charge.

The process of conduction in such a system may be visualized in the following way. An electron moves against an applied electric field by jumping into a vacant state. This transfers the position of the vacant state, or propagates the hole, in the direction of the field, as shown in the diagram. Whether conduction occurs by electrons or holes is determined experimentally from the sign of the Hall emf (*see* HALL EFFECT). If a current is carried in the presence of a magnetic field perpendicular to the current, an emf is developed perpendicular to the current and to the field. The sign of this emf depends on the sign of the charge on the carriers.

Hole-conduction is important in many semiconductors, notably germanium and silicon. The occurrence of hole-conduction in semiconductors can be favored by alloying with a material of lower valence than the "host." Semiconductors in which the conduction is primarily due to holes are called *p* type. Hole conduction is also observed in some metals, including iron and chromium. In other metals, including aluminum and bismuth, both holes and electrons may be present in equilibrium. *See* ELECTRICAL CONDUCTIVITY OF METALS; SEMICONDUCTOR. [J.C.]

Bibliography: C. Kittel, *Introduction to Solid State Physics*, 2d ed., 1956; W. Shockley, *Electrons and Holes in Semiconductors*, 1950.

Holly

The American species of holly, *Ilex opaca*, attains a maximum height of 40–50 ft and has evergreen leaves. It grows naturally in the eastern and south eastern United States close to the Atlantic and Gulf Coasts, in the Mississippi Valley, and westward to Oklahoma and Missouri. It is best known for its bright red berries, which make a pleasing contrast with the deep green spiny leaves, and for this reason it is valued for decorations at the Christmas season. The wood is hard, tough, and close-grained. The heartwood is ivory white when first cut, but becomes brownish with age or on exposure, and takes a high polish. It is used for cabinet work and musical instruments; because it resembles ivory, it is sometimes used for keys for pianos and organs. Its fine grain makes it valuable for wood-engraving work.

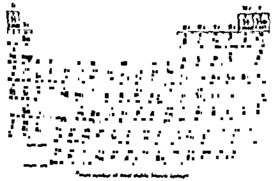
The English holly, *I. aquifolium*, is cultivated extensively in the extreme northwestern United

American holly, *Ilex opaca*. (A. H. Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

States, but is not hardy in the northeastern states. Its spiny leaves are glossier than those of the American holly and have wavier margins. *See* FOREST AND FORESTRY; TREE. [A.H.G.]

Holmium

Element number 67, holmium, Ho, is a metallic element belonging to the rare-earth group. Its atomic weight is 164.94; the stable isotope Ho^{165} makes

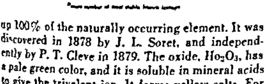


up 100% of the naturally occurring element. It was discovered in 1878 by J. L. Soret, and independently by P. T. Cleve in 1879. The oxide, Ho_2O_3, has a pale green color, and it is soluble in mineral acids to give the trivalent ion. It forms yellow salts. For properties of the metal, *see* RARE-EARTH ELEMENTS.

The metal is paramagnetic, but as the temperature is lowered, it changes to antiferromagnetic and then to the ferromagnetic system. The Néel point occurs about 133°K and the Curie point is in the neighborhood of 20°K. [F.H SP.]

Holocephali

One of two Recent subclasses of the cartilaginous fishes or Chondrichthyes. The Holocephali or chimaeras differ from the other subclass, the Elasmobranchii or sharks and rays, in having only four pairs of gill arches and gills that open to the exterior from a single pair of apertures; in the erectile dorsal fin and spine; naked skin in adults; and in the absence of a cloaca and of ribs. Males are usually equipped with a frontal clasper on the



Deep-water chimaera, *Hydrolagus affinis*; length to 3 ft (After G. B. Goode and T. H. Bean, *Ocean Ichthyology, U.S. Natl. Museum Spec. Bull. 2*, 1895)

head. The teeth are consolidated into six pairs of plates and the upper jaw is immovably fused with the brain case; these adaptations function in grinding mollusks, their chief food. Chimaeras date from the early Mesozoic. They are classified into a single order, the Chimaeriformes, 1 family, the Chimaeridae, 4–5 genera, and about 24 species. All chimaeras are marine, most living in deep water. They are of little economic importance. *See* CHONDRICHTHYES. [R.M.B.]

Bibliography: J. Tee-Van et al. (eds.), *Fishes of the Western North Atlantic*, Sears Foundation for Marine Research, Mem. 1, pt. 2, 1954.

Holometabola

A division of the subclass Pterygota (winged insects) whose members undergo a complete metamorphosis during development. The successive developmental stages are the egg, larva, pupa, and adult or imago. Both larvae and pupae are quite variable in form. The pupa is a quiescent stage and occurs in a protective cocoon or cell. *See* PTERYGOTA. [C.B.C.]

Holothuroidea

A class of Eleutherozoa characterized by a cylindrical body and smooth leathery skin, and known as sea cucumbers. There are no arms but a ring of five or more tentacles may surround the mouth which is usually at one end of the body. There are no pedicellariae. Tube-feet may be present or lacking. There are no ambulacral grooves, although they are represented by internal epineural canals overlying the radial nerves. E. Deichmann (1957) [illegible]

[illegible]

tamerous symmetry is largely concealed by a sec- [illegible]

[illegible]

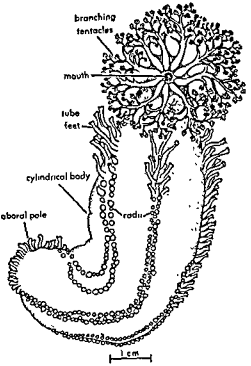
upper (dorsal) surface and a lower (ventral) one. The dorsal side corresponds to the interradius which contains the madreporite, and therefore the ventral side is one of the radii. Each radius runs from the anterior to the posterior end. If tube-feet are developed, their disposition indicates the radii.

The 1100 living species have been grouped in 170 genera arranged in 5 orders: the Elasipoda, Aspidochirota, Dendrochirota, Molpadonia, and Apoda. Species range in size from 3-cm body length up 1.5 m; the largest types are tropical Synaptidae. Colors vary widely; the most brilliant are the Synaptidae. Yellow, red, violet, and fawn tints occur, but many species are somber shades or black *See* APODA (ECHINODERMATA); ECHINODERMATA.

Relation to man. Some holothurians are esteemed as food, in particular about 20 species of the genera *Stichopus* and *Holothuria*. These are fished mainly in the Indian and Pacific Oceans. The animals are boiled for 20 min in sea water, smoked, [illegible]

[illegible]

10,000 tons. In Italy also, fishermen take local spe- [illegible] es, but these are not esteemed. *A few holothurians*



Cucumaria, a representative holothurian.

are poisonous, especially to fish. In the Pacific some species of *Actinopyga* and *Holothuria* yield a secretion used by the islanders as an aid in fishing rock pools, in the same way as rotenone.

Ecology. The Dendrochirota and Apoda have more restricted ranges than the other orders, which are believed to be older. Holothurians occur in all seas, from low-tide level down to the greatest depths explored. The Vitiaz expedition in 1957 took a holothurian from the Tonga Trench at a depth of 6½ miles (10,415 m). At depths below 5½ miles holothurians comprise 90% of the total mass of living matter, the rest being mainly starfishes. Two pelagic genera are known. See ELASIPODA.

Parasites include protozoans, flatworms, nematodes, and annelids. Some crabs, such as *Pinnotheres*, inhabit the cloaca or respiratory trees, and another, *Lissocarcinus*, lurks between the tentacles. Gastropods such as *Entoconcha* and *Entocolax* bore into the skin, body cavity, or foregut. A fish, *Fierasfer*, inhabits the cloaca of large Aspidochirota.

Skeleton. This structure usually comprises no more than a ring of 10 (or 5) calcareous plates around the esophagus and numerous small platelets or spiculae scattered in the skin. In Psolidae the skin plates are large and overlap like scales, but they do not form radial and interradial series. The skin platelets may assume distinctive forms, such as anchors, wheels, or other shapes and may be useful as taxonomic characters.

Muscular system. The muscular system is well-developed and the body can assume a variety of shapes. The chief muscles are five radial longitudinal bands in the body wall; outside these are transverse fibers. Some species have pharyngeal retractor muscles which invert the anterior part of the body. This feature is used as an aid to classification.

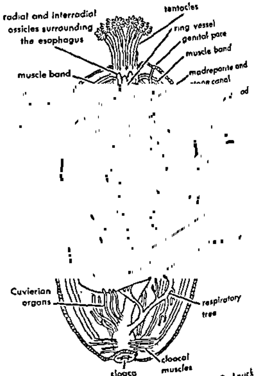
Alimentary system. The viscera lie in the coelom which contains a fluid. The mouth leads into a short esophagus which connects to the long and usually looped intestine. Its front part may be differentiated as a stomach, and the posterior loop serves as a rectum, opening at the cloaca.

Most holothurians are mud swallowers, living on the bacteria and other organic material present in the substrate. The Dendrochirota, however, feed on plankton which is caught on their sticky tentacles and transferred to the mouth.

Some Aspidochirota have paired cloacal glands, the Cuvierian organs, which extrude a mass of sticky threads to trap small animals and to repel predators.

Respiration. Respiration occurs partly in the skin and tube-feet and sometimes in tubular branches of the gut, the respiratory trees. Rhythmic contractions of the cloaca cause sea water to flow to and fro in them.

Hydrocoele. Although poorly developed in the Molpadonia, and vestigial in the Apoda, the water-vascular system is typical in the other three orders. The one or more madreporites lie in the dorsal interradius, internal in nearly all cases, and open into the coelom. The single stone canal which arises from them runs downward to the ring vessel around the esophagus. A single polian vesicle lies in one interradius. The radial vessels run forward to the base of the tentacle ring and then backward in the



Internal anatomy of a holothurian. (*After R. Leuckart*)

body wall, following the radii. The oral tentacles represent specialized tube-feet, and may have ampullae.

Nervous system. This system is typical for the phylum, following the pattern of the water-vascular system, lying between the canals and the body wall. The whole surface of the body is sensitive to light, but some Synaptidae have in addition an eyespot at the base of the tentacles. These are analogous to the eyespots of starfishes and, like them, do not form an image, but are able to detect shadow movements and thus indicate the presence of potential enemies or prey.

Life history. Holothurians seem to attain sexual maturity after about 3 years and continue to grow for several more years. The life span may therefore be longer than in sea stars and sea urchins, perhaps 10 years.

Reproductive system. The reproductive system comprises either one or two gonads, with their ducts. The gonopore is usually anterodorsal. The sexes are usually distinct. Hermaphroditism is common in the Apoda, but individuals are either male or female at any given time, so self-fertilization does not occur. The life history may include a free-swimming larva (auricularia). Many cold-water species incubate their young in various ways. *Leptosynapta minuta* carries the developing embryos in the coelom until they escape by way of the cloaca. Some species of *Psolus* carry the young in dorsal pockets in the skin, covered by calcareous plates. Others shelter the eggs and young by lying on them. A few species, such as *Cucumaria planci*, habitually reproduce asexually by spontaneous transverse fission. Most species can regenerate lost organs. Autoevisceration is a peculiar protective habit which is often displayed in response to interference. The whole gut, the gonads, and respiratory trees are ejected through the cloaca and cast off. Regeneration of the lost structures occurs from the torn mesenteries, and requires several months. During regeneration the animal cannot feed; instead, the muscles are slowly digested in situ. See ASPIDOCHIROTA; DENDROCHIROTA; MOLPADONIA.

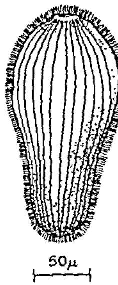
[H.B.F.]

Holotricha

A subclass of the class Ciliata. These protozoa have a fairly uniform body ciliation, as the name implies. Separate articles appear on the groups listed in the following classification:

Subclass Holotricha
- Order Gymnostomatida
  - Trichostomatida
  - Chonotrichida
  - Hymenostomatida
  - Thigmotrichida
  - Peritrichida
  - Suctorida
  - Apostomatida
  - Astomatida

The cilia are typically arranged in longitudinal



Prorodon, a primitive holotrich

rows over the body, although scattered exceptions exist. Cirri are absent and buccal membranelles are absent or inconspicuous, except in one group. A mouth is often, although not always, present. Ciliary organelles associated with it are generally not conspicuous. Trichocysts are present in species belonging to a number of families. The nine orders contain several thousand species and include groups of, presumably, the most primitive ciliates living today. The prototype of the Holotricha is exemplified by the form portrayed in the illustration.

[J.O.C.]

Homeostasis

The essence of living substance is that it differs chemically from the surrounding medium and yet maintains a dynamic equilibrium with its environment. Homeostasis is the maintenance of internal constancy and independence of the environment. It occurs both at the cellular level and in the fluids and organ systems of multicellular organisms. Referring to homeostasis in mammals, the nineteenth-century French physiologist Claude Bernard coined

physiologist Walter B. Cannon, who first used the word "homeostasis," referred to the systems of checks and balances which maintain internal constancy as "the wisdom of the body." Homeostasis is usually considered as occurring in animals although the term may apply as well to plants.

The cells of multicellular organisms are bathed in fluid which is kept relatively constant in its ionic composition, osmotic concentration, pH, level of sugar, and organic composition. No organism could be as dilute as fresh water and survive; terrestrial organisms must be protected against desiccation. The closest approach to conformity with the medium is in some endoparasitic worms and protozoans and in some marine invertebrates.

Patterns of adaptive responses. Two patterns of adaptive responses of animals to environmental change are recognized. (1) An animal may alter a given property with the environment; it conforms to the medium and homeostasis consists in cellular adjustment such that metabolism continues in an altered state. For example, in cold-blooded animals the temperature conforms to that of the environment, yet activity may be high. (2) An animal may regulate its internal state and maintain internal constancy despite an altered environment. Such regulation is by a series of automatic feedback controls as environmental stress is applied, and at some environmental limits regulation fails and the animal cannot long survive. For example, warm-blooded animals maintain a constant body temperature over a range limited by both heat and cold.

The cellular level. Homeostasis at the cellular level is universal in living organisms. The first aggregates which could be called living cells must have been separated from their marine environment by a bounding layer which prevented free interchange. The cells of all plants and animals differ chemically from extracellular fluid. An important mechanism of cellular individuation resides in the cell surface, which has selective permeability and which, in many cells, provides some mechanical rigidity. The cell surface consists of a relatively inert pellicle plus a plasma membrane of protein and lipid which permits entry and exit of relatively few kinds of organic molecule and which varies in its permeability to inorganic ions according to activity. The surface also has some enzymes or carriers for active transport. Intracellular homeostasis consists of regulation of metabolism in its broadest meaning, that is, oxidation of foodstuff at such a level that sufficient energy will be available for necessary work and yet not at such a pace as to burn excessively. Many enzymes are synthesized according to substrate load, and alternate metabolic pathways may be used according to steady-state conditions existing between a cell and its environment. See CELL MEMBRANES AND MONOLAYERS.

Homeostatic mechanisms. A catalog of the mechanisms of homeostasis would be a textbook of the physiology of organ systems and of cells. Morphological changes may occur in response to environmental stress. Arctic birds and mammals tend to increase their coats of feathers or fur and insulating fat in winter, tadpoles grow larger gills when reared in low oxygen, men living at high altitudes have increased concentrations of blood hemoglobin, and bone structure varies according to mechanical stress. The sequence of homeostatic responses to a severe sudden injury-type stress varies according to the kind of animal. H. Selye has described, for mammals, the nonspecific "general adaptation syndrome" which consists of an initial alarm reaction of a shock phase and then a countershock during which, under pituitary activation, the adrenal cortex enlarges, lymphocytes decrease in the blood, and the shock reactions disappear. Resistance then develops, sometimes followed by a stage of exhaustion.

Sequential reactions. The sequence of responses to cold illustrates some of the principles of sequential reactions in complex homeostasis. First, cold receptors in the skin are stimulated, which results in constriction of peripheral blood vessels, and erection of hairs. If the body is then chilled, the temperature-sensing portion of the hypothalamus is stimulated, shivering begins, and metabolism increases, partly under endocrine control. The first line of defense is insulative, the second is increased heat production. The complex reactions of fear, rage, courtship, and care of young require a variety of sensory and effector structures integrated by circulatory, nervous, and endocrine systems. Alternate means of attaining the same end are well known. Epinephrine, the secretion from the adrenal medulla, has physiological action similar to impulses in the sympathetic nervous system. The digestive system has a sequence of proteolytic enzymes, overlapping in function yet somewhat specific. The hierarchy of controls is noted, particularly for endocrines and nervous system. The anterior pituitary regulates the secretions of adrenal cortex, thyroid, pancreas, and other glands; the posterior pituitary regulates kidney function. Another principle is that of checks and balances. The vagus nerves slow the heart and stimulate the intestine, while the sympathetics accelerate the heart and relax the intestine. When motor nerve centers for one set of muscles are active, those for antagonistic muscles are inhibited. Various regions of the brain counterbalance other regions. Many sense organs show spontaneous activity which is modulated by stimulation, as either an increase or decrease in signals to the central nervous system. Many "conforming" animals compensate for environmental change with the result that relative internal constancy is maintained. Most aquatic poikilotherms, for example, fish, acclimated to cold, have a higher oxygen consumption than individuals acclimated to warmth, when both are measured at an intermediate temperature. Animal behavior is often compensatory for environmental change. See BEHAVIOR, ONTOGENY OF.

In individual cells, homeostatic mechanisms are less well known. The presence of substrate induces the synthesis of enzymes for attacking it. The presence of products of substrate breakdown may retard specific enzyme formation. Entry and exit of materials across cell surfaces is controlled by intracellular concentrations.

In all the systems of homeostasis, the hierarchies of controls, the checks and balances, the alternate mechanisms, the compensatory reactions, there are marked analogies to servo systems. Sensing elements are poised at critical levels, and upon deviation, feedback mechanisms are brought into action and tend to restore equilibrium. [C.L.P.]

Bibliography: W. B. Cannon, *The Wisdom of the Body*, rev. ed., 1939; C. L. Prosser, Physiological variation in animals, *Biol. Revs.*, 30:229, 1955.

Homobasidiomycetidae

A subclass of basidiomycetous fungi (Basidiomycetes), also known as Holobasidiomycetes or Eubasidiomycetes. In this subclass the basidium is not divided by cross walls. There [illegible] of Homo[illegible] the Gaste[illegible] include the mushrooms, pore fungi, and their relatives, the basidia are formed in an exposed layer (hymenium), and the basidiospores borne asymmetrically on slender stalks (sterigmata) are forcibly discharged. In the Gasteromycetes, which include the puffballs, earthstars, stinkhorns, and their relatives, the basidia are enclosed during their formation, and the basidiospores borne symmetrically on long sterigmata are not forcibly discharged. *See* BASIDIOMYCETES; MUSHROOM.

[R.M.P.]

Homoptera

One of the two suborders of the Hemiptera. This is a major group of sucking insects, with more than 30,000 species, even though, in Asia and Africa, the number of undiscovered species probably still exceeds the discovered ones. Common examples are the cicadas, aphids and leaf hoppers. The suborder is difficult to characterize because of the large number and diverse forms of the species it contains. Because of this, many authorities consider the Homoptera to be an order. The head of these insects is so hypognathous that the beak appears to arise from the ventral posterior margin of the head or even from the prosternum. The gula is membranous or absent. As in the suborder Heteroptera, the beak consists of two pairs of stylets, formed by the maxillae and the mandibles, ensheathed in the labium. The maxillary stylets fit together to form a double tube, one channel serving for the passage of food and the other for saliva.

Most winged species have four wings, but male scale insects have only two. In most forms, both pairs of wings are membranous and transparent, but in some, the forewings are somewhat thickened and may then be either coriaceous and translucent or opaque, and with or without an apical membranous area. When the insects are at rest, the forewings are usually held, rooflike, over the dorsum, with the apex of one of them slightly overlapping the apex of its complement (Fig. 1).

The digestive tract is peculiarly complex in a vast majority of species, in that it forms a filter chamber, a structure consisting essentially of a close association of the posterior end of the midgut with the posterior end of the foregut. The approximated portions of the loop thus formed are believed to permit certain elements of the food to bypass most of the digestive portion of the gut. This anatomical feature, coupled with the prodigal feeding habits of most of the species, has led to the theory that certain factors in plant sap are present in such small quantities that large amounts of the sap must be imbibed, and the unneeded component shunted across the loop of the filter chamber, in order to obtain a sufficient quantity of the factors needed to sustain the insect. All species for which the food habits are known are phytophagous, deriving their nourishment from plant sap.

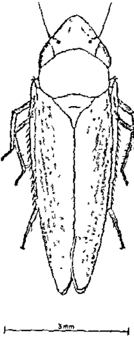


Fig. 1. A cicadellid, dorsal view.

In most species metamorphosis is gradual, but in a few it is practically holometabolous. The adults and nymphs of most species are terrestrial, but a few species are subterranean in all stages and others are subterranean only in the immature stages. A number of species are vectors of virus diseases of plants.

Series Coleorhyncha. This group is characterized by the origin of the beak, formed at the anteroventral extremity of the face, and by the fact that the propleura form a sheath for the base of the beak. The hindwings are absent, and the forewings are held flat over the abdomen in repose. The flight function has been lost. The prothorax is provided with dorsolateral expansions, the paranota, absent in other modern insects, but similar to structures found in Permocarboniferous fossil insects. There are additional anatomical features which have been cited as evidence of primitiveness. These serve as a basis for placing the Peloridiidae, the only known family, in the lowest position among the Homoptera. The species lack a filter chamber in the digestive tract, and are rare and few in number. They occur in Tasmania, New Zealand, and South America.

Series Auchenorhyncha. This series and the Sternorrhyncha are the major groups of the Homoptera. In the *Auchenorhyncha*, the beak arises at the anteroventral extremity of the face, and is not sheathed by the propleura. The labium arises

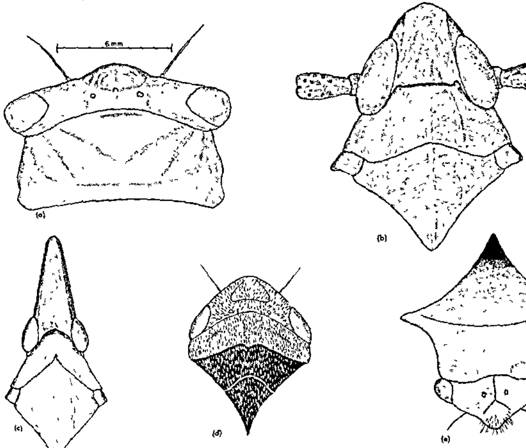


Fig. 2. (a) Anterior dorsum of a cicada. (b) Anterior dorsum of a fulgoroid (apical part of antenna omitted in one). (c) Anterior dorsum of another species of fulgoroid. (d) Anterior dorsum of a cercopid. (e) A membracid, anterior view of face and pronotum.

well in front of the anterior legs. The antennae usually have one to three basal segments surmounted by a seta. In repose, the forewings are usually placed in a rooflike manner over the abdomen. They are more heavily sclerotized than the hindwings in some species, and may have a terminal membrane, thus somewhat resembling the forewings of the Heteroptera. Many species are good flyers. The hindwings nearly always have a jugum. The tarsi are nearly always three-segmented in the adults. Hindlegs are often adapted for jumping.

A filter chamber is apparently a constant feature of the digestive tract, although it may be greatly reduced in some species. There are usually four Malpighian tubules, originating in the anterior portion of the hindgut in the filter chamber. These frequently have an enlarged portion to which a glandular function has been attributed. Some species are [illegible] tion by the [illegible] to be of wi [illegible] rhyncha and probably occurs throughout the group. A sound-producing apparatus has also been demonstrated in the females of some species.

The Auchenorhyncha includes a large number of species. A number of classifications have been proposed, most of them differing chiefly in the rank assigned to the higher categories. The classification adopted here is a common one. It divides the series into the superfamily Fulgoroidea and the families Cicadidae, Cercopidae, Membracidae, and Cicadellidae. These families are not subordinate to the superfamily Fulgoroidea.

Superfamily Fulgoroidea. These are insects commonly known as the lantern flies. They are distinguished from other Auchenorhyncha by the following characteristics (Fig. 2*b*, *c*). The middle coxae are the same length as the anterior coxae and are joined to the body at some distance from the median line. Tegulae, small, scalelike sclerites, are usually present at the base of the forewings. The hindwings usually lack a submarginal vein parallel to the wing margins.

The antennae, situated beneath the eyes, have two well-developed basal segments, of which the second is enlarged and often provided with numerous sensillae. Longitudinal carinae are often found on the face. There may be two or three ocelli, also located on the face. The forewings may be tectiform, vertical, or horizontal with their apices overlapping in repose. The anal, or claval, veins are usually fused to form a Y-shaped pattern

There is no submarginal vein in the hindwing. Contradictory statements have been made regarding the presence of a filter chamber in the digestive tract.

The popular name, lantern flies, resulted from old reports of bioluminescence in the head of a large South American species. This phenomenon has not been observed since, but the common name has survived.

This group is subdivided into 20 families and includes many economic species which are important because of the damage they do while feeding or because they carry virus diseases of plants.

Family Cicadidae. Included in this group are the cicadas, harvest flies, and jar flies. The insects in this family are probably better known to the layman than any other homopterous family. They have received this attention because of their large size and the strident songs of the males. The following combination of characteristics will separate them from other families of Auchenorhyncha. The short middle coxae differ from the anterior coxae and are joined to the body near the median line. Tegulae are absent. The hindwings have a submarginal vein parallel to the wing margin. There are three ocelli arranged in a small triangle (Fig. 2*a*). The immature forms have digging legs.

The head of the adult is large, with protuberant eyes (Fig. 2*a*). The upper median portion (postclypeus) of the face is swollen, and this area is bounded throughout its length by the lora, or mandibular plates. The anterior and posterior tentorial arms are connected in the head. The forewings are membranous, tectiform in repose, having a single anal vein or two more or less parallel anal veins. The anterior legs have dilated femora which are spiny beneath. The posterior legs are not saltatorial. In the digestive tract, there is a long filter chamber in which the two parts of the gut are spirally interwound, with the hindgut issuing from its anterior end. There are four Malpighian tubules.

The adults are usually found on trees or shrubs. At least in some species, the songs of the males assemble local populations. After mating, the female lays eggs in the shoots of plants, making a slitlike incision for them by means of an ovipositor which is provided with sawlike blades or valves. These incisions frequently result in the death of the shoots. After hatching, the young cicadas fall to the ground and take up a subterranean existence which may last from 1 to 17 years. During this period they feed on plant roots. The North American 17-year cicada, *Magicicada septendecim* L., a species with a 17-year life cycle in the northern part of its range, and a 13-year cycle in the southern part, and with a number of broods in each part, attracts much attention because of its occasional occurrence in tremendous numbers. It is frequently referred to, in popular accounts, as the 17-year locust, possibly as a result of an association by early North Americans with the biblical plagues of locusts. There is no basis for such an association, other than large numbers of individuals, and the use of the term locusts to refer to cicadas is inadvisable. See POPULATION DYNAMICS.

Family Cercopidae. The spittle bugs and froghoppers are common examples of this group. The insects in this group most often attract attention in the immature stages, during which they surround themselves with a mass of froth or spittle. The family has the following distinguishing characters. The middle coxae, the tegulae, and the hindwings are as described for the Cicadidae. There are usually two ocelli, never three, occasionally none. When present they are located on the crown of the head, of which the median apical portion is usually distinctly delimited from the remainder by sulci (Fig. 2*d*).

The head of the adult is proportionately smaller than in the Cicadidae, but the structure of the tentorium is similar. The pronotum, which does not extend posteriorly, is horizontal or sloped, but not vertical. The forewings, tectiform in repose, are usually somewhat sclerotized and different in texture from the membranous hindwings. The hind legs are saltatorial, with the coxae short and conical, not transversely dilated. The tibiae have one or

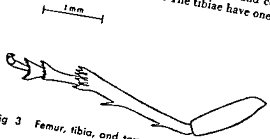


Fig 3 Femur, tibia, and tarsus of a cercopid.

a few stout spurs and a cluster of small outgrowths at the apex (Fig. 3). In the digestive tract, there is a complicated filter chamber in which a number of convolutions are found. The hindgut issues from its posterior end. There are four cryptonephridial Malpighian tubules arising from two basal stalks.

The froth, or spittle, which covers the immature insects consists of extruded anal fluid with which air bubbles are mixed. The frothy mass does not evaporate readily, a fact which seems correlated with the reported inability of the young cercopids to survive in a dry atmosphere. There are many more species in the tropics than in temperate climates. One species, *Philaenus leucophthalmus* L., the meadow spittle bug, is so common in the temperate portion of the Northern Hemisphere that its masses of spittle are familiar sights to almost everyone. This species has been shown to be an efficient vector of a "yellows" virus of peaches. Some species which live in the Australian region and the East Indies live within a calcium carbonate tube attached to stems or leaves.

Family Membracidae. The treehoppers are small to medium in size and seldom attract attention. Most of them feed on woody plants and are found on the stems in sunny locations. Frequently, a num-

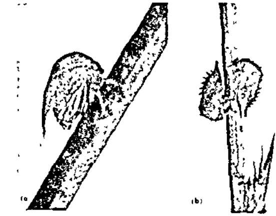


Fig. 4 (a) A membracid adult on a stem. (b) Membracid nymphs on stems (Courtesy of C. H. Hanson)

ber of specimens are arranged in a vertical row on the stem, all with their heads downward Membracidae may be recognized by the following combination of characters. The middle coxae and the tegulae are as described above for the Cicadidae. There are two ocelli, or none. The pronotum extends backward over the abdomen, sometimes almost completely covering the wings (Fig 4a). The upper portion of the head is vertical (Fig. 2e).

In the adult head, the anterior and posterior arms of the tentorium are not connected. The antennae consist of two larger basal segments and a terminal seta with a large number of subsegments. The forewings are usually membranous but may be partly sclerotized, and are tectiform in repose. The radial and medial veins are not fused basally. There is often a submarginal vein in the hindwings. The anterior and middle tibiae are often dilated. The hind legs are saltatorial and spiny. The digestive tract has a filter chamber with fewer coils than found in many Homoptera. There may be two or four Malpighian tubules which are cryptonephridial, at least in some species.

The greatest number of species occurs in the warmer regions of the world In many species the enlarged pronotum has adornments, excrescences, and processes which are astonishing in appearance, some of them nearly as large as the remainder of the insect. Most species are not active, leaping only to take flight, seldom flying, and then only for short distances. In North America, *Ceresa bubalus* (Fabr.), the buffalo treehopper, is an economic species which damages fruit and other trees by slitting the twigs during oviposition. The nymphs (Fig. 4b) leave the trees and feed on herbaceous plants, often occurring in great numbers in pastures. The adults return to woody plants before oviposition.

Family Cicadellidae. The leafhoppers are included in this large family. These usually small insects are known to many people by sight but not by name, because of their common occurrence in great numbers at night near lights. The species may be distinguished from other families by the following combination of characters. The coxae and tegulae are as described for Cicadidae. The pronotum does not extend backward over the abdomen and does not have a median ridge. The upper portion of the head is never vertical. There is usually a submarginal vein parallel to the wing margin in the hind wings (Fig. 5). There are two ocelli, or none. When present, they may occur on the face, on the crown of the head, or on the margin between the face and crown. The hind tibiae have many spines arranged in rows (Fig. 1).

On the face, the lora border the postclypeus for only a short distance. The anterior and posterior arms of the tentorium are not connected in the head. The forewings may be membranous or heavily sclerotized and, in the latter case, often have a membranous apical portion. They are usually tectiform in repose. Radial and medial veins arise from a common stalk. The hind legs are saltatorial. The digestive tract has a filter chamber, but reports indicate a considerable degree of variation in its detailed structure. There are four cryptonephridial Malpighian tubules.

Probably the greatest number of species occurs in tropical areas, but the majority of these have not been described. Temperate North America also has a large number of species. Leafhoppers occasionally bite man but apparently they have never been seen taking blood. In recent years, several species have been found to be vectors of virus diseases of plants. Some of the more important of these virus diseases are phloem necrosis of elms, curly top virus of sugar beets, aster yellows virus, rice dwarf virus, and papaya bunchy top virus, in addition to virus diseases of potatoes, clover, alfalfa, grapes, peaches, sugar cane, eggplant, corn, maize, wheat, and cranberries. In a few cases, the virus is transmitted transovarially. See PLANT DISEASE.

... of homop ...

are usually long, filamentous, and have no well differentiated terminal seta. Wingless forms are common. The wings, when present, are usually membranous, with reduced venation, and usually without closed cells. The hindwings have no jugum The tarsi of the adults are 1- or 2-segmented A filter chamber is usually present in the digestive tract.

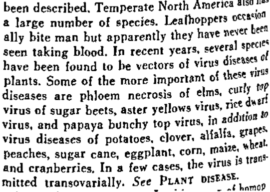


Fig. 5. A cicadellid wing. The more basal broken line represents the jugal fold.

The Sternorrhyncha includes a large number of species, many of them of great economic importance. The winged forms are not strong flyers, but they are so light that they may be borne considerable distances by air currents. There have been several classifications proposed for the Sternorrhyncha. The following subdivision into superfamilies and families appears to be the most widely used classification.

Family Psyllidae. The Psyllidae are known as the jumping plant lice. This family has also been known as the Chermidae. Its representatives resemble cicadas in appearance, but are much smaller. About 1000 species are known. The adult insects have a transverse head, usually emarginate anteri-

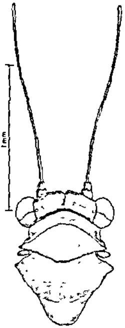


Fig 6. Anterior dorsum of a psyllid.

orly, with protuberant eyes (Fig. 6). The face is often in the form of two conspicuous cones which are directed ventrally. Three ocelli occur. The antennae are 6- to 10-segmented, with the latter condition being common. They consist of two wider basal joints and a terminal portion which bears sensillae. The apical antennal segment bears two fine setae. The winged forms have four wings which are usually membranous, but the forewings are thicker than the hindwings. The wings are tectiform in repose. The venation is reduced, but conspicuous, and cross veins are seldom present. In the forewings, the radius, media, and cubitus arise from a common stalk. The tarsi are 2-segmented with a pair of apical claws. The hind legs of the adult are saltatorial, with the coxae greatly dilated.

The hindgut and the posterior portion of the esophagus are looped around each other. Possibly this is a functional filter chamber, although a specialized clearly delimited filter chamber, similar to that found in some other groups of Homoptera, is apparently absent. There are four Malpighian tubules opening separately into the midgut.

The species of psyllids may be monophagous or polyphagous. There may be an alternation of host plants, but the nymphs and adults occur on the same plant. Hibernation may occur in the egg, nymphal, or adult stages. The nymphs have flattened bodies and do not have saltatorial hind legs. They are frequently covered by a secretion which may be wooly, waxy, or viscous.

Some psyllids produce severe damage to their food plants. This damage may result from the mere feeding by tremendous numbers of individuals, from the resulting yellowing or rolling of leaves, or from galls produced on the leaves. Indirect damage may result from the growth of fungi on leaves which have become coated with the sugary excrement, the honeydew, of the psyllids. The pear psylla, *Psylla pyricola* Forster, is a species of considerable economic importance in North America. Its damage includes premature falling of leaves and fruit, reduced quality of fruit, and blackening of the leaves from fungus.

Family Aleyrodidae. The whiteflies are 7 mm or less in length and usually lightly covered with a white, powdery, waxy material which has led to their common name. The head varies in shape, and there are two ocelli. The antennae are usually 7-segmented, with the two basal segments thicker than the other segments which may have sensoria and terminate in a bristle. The eyes may be reni-

and there are no cross veins. The tarsi are 2-segmented, with a pair of terminal claws between which is a median structure which varies in form from a blade to a bristle. The thorax is separated from the abdomen by a constriction. There is a filter chamber in the gut, but the first ventriculus of the gut is not greatly dilated where the filter chamber occurs. There are only two Malpighian tubules. *The stylets of the mouthparts, when not in use, are*

newly-hatched insects are ambulatory, but they soon choose a feeding site and remain virtually sessile until reaching the adult stage. The sessile forms, with their appendages greatly reduced, resemble scale insects and the last larval instar is anchored to the plant by a waxy secretion. Both bisexual reproduction and parthenogenesis occur within the group.

Whiteflies directly damage plants from their feeding. Indirectly, damage results from spotting at the feeding site, growth of fungus on the excreted honeydew, or from increased susceptibility of leaves to winter damage. Two species are important pests in the United States. *Dialeurodes citri* (Ashmead), the citrus whitefly, an Asian species, is an important pest of citrus and other plants

in California, Florida, and elsewhere. *Trialeurodes vaporarium* (Westwood), the greenhouse whitefly, is an important pest in greenhouses and on several cultivated crops. Other species carry virus diseases to cassava, cotton, cucumber, sunflower, and tobacco.

Superfamily Aphoidea. In this large superfamily of small insects there are four wings, or none. The wings, usually membranous or whitish and opaque, are held tectiform over the abdomen in repose. The forewings are much larger than the hindwings. The tarsi are 2-segmented and usually have two claws. The hind legs ordinarily are not saltatorial. The intersegmental lines are distinct in the abdomen. Both parthenogenetic females and sexual females occur. The life history may be extremely complicated, involving several host plants as well as several forms of the insects. Honeydew is usually produced. This superfamily includes the families Aphidae and Chermidae.

Family Aphidae. The true aphids (Fig. 7*a*) are members of this family in which the sexual females are oviparous and the parthenogenetic forms are viviparous. The sexual females, and usually the males, have a functional beak and a continuous digestive tract. Cornicles are usually present.

Compound eyes are present in the adults and three ocelli occur in the winged forms. Usually the antennae are 5- or 6-segmented with the apical segment ending in a tapering terminal process. Sensoria are present on some antennal segments. The beak has five segments. The wings extend caudad beyond the apex of the abdomen. In the forewing, vein Rs (radial sector) separates from the stigma and reaches the wing apex. The last abdominal segment is extended to form a cauda, with the anus opening beneath it. A poorly developed filter chamber is said to occur in the gut of some aphids, but a number of aphids lack it. Malpighian tubules are absent throughout the family (Fig. 8).

Many species of aphids have both an alternation of generations and an alternation of hosts, one of which is usually obligatory (*see* METAGENESIS). In temperate climates, the eggs are laid most commonly on a woody host in autumn where they overwinter. The females which hatch from the eggs produce, by viviparous parthenogenesis, a variable number of generations on the primary host. Winged forms are eventually produced which migrate to a herbaceous alternate host. Here a number of generations are produced by viviparous parthenogenesis, some of the individuals having wings, but most of them being wingless. In autumn, a generation of winged females occurs which migrate to the primary host. Here they produce parthenogenetically the winged sexual males and females. After mating, the females lay eggs. There are many departures from this generalized life history. Some species are monophagous. In others, alternation of hosts appears optional, the aphid species being able to reproduce on the primary host, secondary host, or both.

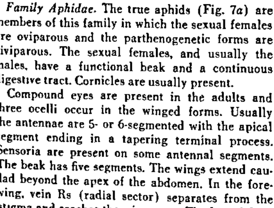


Fig. 7. (a) A colony of aphids on a rose shoot (C. F. Smith). (b) The citrus mealybug (Ohio Agricultural Experiment Station).

Many species of aphids, because of the honeydew they produce, are very attractive to ants, and colonies of very small aphid species, which otherwise might escape notice, can often be located by observing the attending ants (*see* SOCIAL INSECTS). In a few species, this relationship has progressed to the point where the ants are necessary for the survival of the aphid species, as in the corn root aphid, *Anuraphis maidi-radicis* (Forbes). The eggs of this species are cared for by the corn field ant, and the young aphids are transferred to corn roots early in the growing season.

Many aphid species are important because of damage done in feeding. The pea aphid, *Macrosiphum pisi* (Harris), is an important pest to peas and other crops in the United States. The cotton aphid, *Aphis gossypii* Glover, is an important pest of cotton, melons, and other crops in the same region. Other species are important because of their ability to transmit virus diseases to plants. The plants involved include potatoes, sandalwood, squash, beets, cauliflower, celery, clover, cucumber, groundnut, alfalfa, onion, pea, sugarcane, tobacco, turnip, and citrus.

Family Chermidae. This is a small family of minute insects, the adelgids and phylloxeriids, in which both the sexual and parthenogenetic females are oviparous. Cornicles are absent. The antennae are 3- to 5-segmented and bear sensilla. Both winged and wingless forms occur. The following are subfamilies of Chermidae.

In the subfamily Cherminae all forms have a beak and the digestive tract is not closed; the wings are tectiform in repose, and there is no branched vein extending toward the posterior margin of the forewing. There is frequently a waxy, flocculent secretion. The life cycle is extremely complicated and may involve an alternation of plant species and an alternation of generations in the insects, with several morphological forms occurring in the life history of one species. The primary host is always spruce. Secondary hosts are other conifers.

In the subfamily Phylloxerinae, the sexual forms lack mouthparts. The parthenogenetic females have

a beak but the digestive system is closed (not continuous), and there is no honeydew. In repose, the membranous wings lie flat over the abdomen. There is a branched vein extending towards the posterior margin of the forewing. Waxy secretions, rarely present, are not flocculent. Although winged migrants occur commonly, there is no secondary host.

The grape phylloxera, *Phylloxera vitifoliae* (Fitch), has been a severe pest of cultivated grapes and once threatened the entire wine industry of France. It was discovered that native North American grape roots were not damaged greatly by the phylloxera, and consequently these were used as grafting stock in Europe to reduce damage.

Superfamily Coccoidea. The scale insects and mealy bugs, which are members of this large and important superfamily, are usually small. More than 4000 species have been described. In the males, the hindwings are reduced to clublike halteres. The wings are usually held flat over the back in repose, and the venation is greatly reduced. The females are wingless. When legs are present the tarsi are usually 1-segmented, but in some males the tarsi are 2-segmented. There is a single tarsal claw. The hind legs are not saltatorial. In some species, the abdominal segmentation is much modified or obliterated.

The males are usually very small, even in species where the female is much larger. They have no beak, do not eat, and have a nonfunctional digestive tract. Young males resemble corresponding female stages, but they molt more often, eventually passing through a stage so much like the pupal stage of holometabolous insects that they have presented an obstacle to systematists of higher categories of insects who used the type of metamorphosis as a fundamental criterion in classification. Adult males normally have long antennae consisting of 10 or fewer segments, although written accounts have stated, as a result of erroneous observations, that as many as 25 segments were present. The eyes of the adult male are compound in some species, but they usually consist of a series of isolated facets. The legs are well developed. The caudal end of the abdomen often bears an elongate process. Males are unknown in some species, and apparently of rare occurrence in some others. A few species have wingless males.

The females are usually sedentary, with the legs absent or reduced and nonfunctional. The body may be soft, or gall-like, or covered with powdery or tufted wax, or with scales or other hardened secretions. Some species form cysts which resemble small pearls. Others inhabit plant galls. In some groups, the females superficially resemble aphids, but cornicles are never present. The females usually have a 1- to 3-jointed beak which arises behind the bases of the anterior legs and varies considerably in shape and size. Its stylets are coiled in a ventrally located internal crumena when not in use. Antennae are present in adult females and may be reduced to flat disks, or may be elongate, with as many as 11 joints. The eyes are very simple and somewhat resemble ocelli. There are usually a number of glands, pores, and ducts in the integument and these features are useful in taxonomy. There is no pupalike stage, as occurs in males. The gut may be either continuous, and with a complex filter chamber, or discontinuous. There are two, three, or four Malpighian tubules.

In the first nymphal stadium, both sexes are mobile and known as crawlers. At this period in their development, they spread over the plant and to other plants.

Scale insects injure the host plant by their feeding on leaves, stems, or roots, and a number of species are very important economically. Most species produce large quantities of honeydew. A few species have been shown to be vectors of virus diseases of cucumbers, tobacco, and cacao. The sedentary habits and small size of the females have contributed to their wide distribution on economically important plants. Monophagous and polyphagous species occur.

Although specialists in scale insects generally agree that the included groups of insects form a superfamily, there seems to be little agreement on the constitution of the several included families. For this reason the taxa discussed here can be treated most conveniently, at present, as subfamilies. As many as 20 taxa in the family group have been accepted by some authors.

In the Margarodinae, abdominal spiracles are present in all stages of development. The adult males have compound eyes, and usually 10-jointed antennae. Adult females have at least some segmentation of the legs, and usually two bristles on the tarsal claw. The antennae of the adult female are not contiguous basally, and the [illegible] underground and have become lawn pests in some parts of southern United States in recent years, usually in areas where the soil is sandy. The females form glassy cysts in which they are able to live quiescent for a considerable time without food. The cysts, or "ground pearls," are lustrous and are [illegible] seta on the terminal segment. The adult males have a strongly bivalved penis sheath. The body of the female is covered with hard white waxy plates. *Orthezia* [illegible] anus is dorsal. This subfamily includes the cottony-cushion scale, *Icerya purchasi* Maskell, a species introduced to California from Australia and which at one time threatened the citrus industry in California. It was brought under control by an introduced Australian coccinellid, *Rodolia cardinal*[illegible]

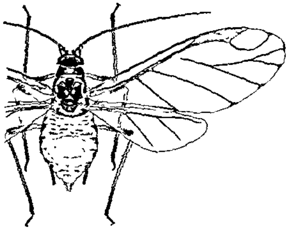


Fig. 8. An aphid, dorsal aspect.

Mulsant, in one of the earliest known instances of biological control.

In the Aspidiotinae, the adult male does not have compound eyes. Adult females have greatly reduced antennae and are without legs. There are no abdominal spiracles in any of the stages. There are no setae at the anal opening. In the females and nymphs, the apical abdominal segments are fused to form a compound pygidium. The scales are shieldlike and nearly circular. The San Jose scale, *Aspidiotus perniciosus* Comstock, is one of the most [illegible] and [illegible] its [illegible] were [illegible] in the United States. It attacks many kinds of fruit trees and shrubs, as well as woody ornamentals. It occurs throughout United States and Canada and in other parts of the world in temperate regions.

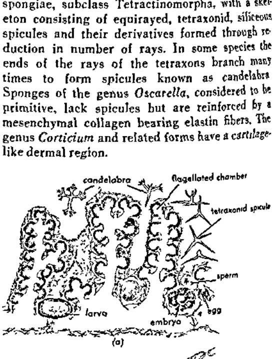
The species of Lepidosaphinae have dark-colored scales and features similar to those of the Aspidiotinae, except that the scale is not circular. Three species of *Lepidosaphes* are of economic importance, two on citrus and one, the oyster shell scale, a cosmopolitan species which infests many kinds of woody plants.

In the Lacciferinae, the adult male does not have compound eyes. The abdomen is without spiracles in all stages, and in the nymphs and females its apical segments are not coalesced to form a pygidium. The abdominal apex has a tubular projection. The body is not covered by a scale, but the insects are enclosed in a mass of resin. The Oriental species, *Laccifer lacca* (Kerr), or lac insect, found on a number of host plants, as well as related species, secretes the resinous material from which shellac is produced (*see* SHELLAC).

The Eriococcinae have many characteristics like those of the Lacciferinae, but the abdominal apex is without the long tubular projection. The adult females and late instar nymphs have an anal ring. The beneficial cochineal insects, *Dactylopius coccus* Costa (*see* ANTHRAQUINONE PIGMENTS), belong to this group. The harmful mealybugs, species of *Pseudococcus*, (Fig. 7*b*), and other genera also belong here. These soft-bodied insects are covered with flocculent wax secretion which suggested their common name. A number of species are harmful to cultivated plants, and some are serious pests of greenhouse and house plants. *See* ENTOMOLOGY, ECONOMIC; HEMIPTERA; INSECT PHYSIOLOGY; INSECTA. [D.A.T.]

Homoscleropharida

An order of primitive sponges of the class Demospongiae, subclass Tetractinomorpha, with a skeleton consisting of equirayed, tetraxonid, siliceous spicules and their derivatives formed through reduction in number of rays. In some species the ends of the rays of the tetraxons branch many times to form spicules known as candelabra. Sponges of the genus *Oscarella*, considered to be primitive, lack spicules but are reinforced by a mesenchymal collagen bearing elastin fibers. The genus *Corticium* and related forms have a cartilage-like dermal region.



(*a*) Section through *Plakina*. (*b*) Section through *Plakortis* with cortex. (*From Hyman, 1940, after Schulze, 1880*)

Homoscleropharid sponges are mostly small in size and encrusting to massive in shape. They occur in tidal and shallow waters, down to depths of at least 500 meters. Fossil sponges with spicules suggesting homoscleropharid affinities are scattered through the fossil record from Carboniferous strata upward. *See* DEMOSPONGIAE; TETRACTINOMORPHA. [W.D.H.]

Honeydew melon

A long-keeping variety of muskmelon, *Cucumis melo*, of the plant order Campanulales. The fruit is large (5–7 lb), oval, smooth, creamy yellow, and without surface markings. The "flesh" is thick

(1¼-2 in.), light green, juicy, sweet, with very mild aroma.

Honeydew is an American name for the variety White Antibes which was grown in France and Algeria long before its introduction into the United States. The honeydew melon requires a warm season of about 125 days. It is very susceptible to diseases, which are intensified by rain or high humidity. Except in California, Texas, and Colorado, it is little grown in the United States. The average annual farm value in the U.S. for the period 1949-57 was about $6,750,000. See CAMPANULALES; MELON GROWING; MUSKMELON.

[V.R.B.]

Honing

The process of removing a relatively small amount of material from a surface by means of abrasive stones to obtain a desired finish or extremely close dimensional tolerance. Seldom are more than a few thousandths of an inch of stock removed. The abrading action of the fine grit stones occurs on a wide surface area rather than on a line of contact as in grinding. As it applies to machining, honing refers primarily to work done on cylindrical surfaces. In addition to metals and carbides, materials such as plastics, ceramics, and glass may be honed.

A hone consists of a holding device containing several oblong stones arranged in a circular pattern. These may be set at a given diameter or forced against the work by a wedging action. The hone floats in the hole as it rotates and reciprocates. While it will remove high spots and surface inaccuracies, it will not correct the position of a hole or establish its alignment. For external honing, the workpiece usually rotates and reciprocates.

Honing is done with manually operated equipment as well as with vertical and horizontal honing machines. See GRINDING; MACHINING OPERATIONS

[A.T.]

Hooke's law

A generalization applicable to all solid materials, stating that stress is directly proportional to strain and expressed as

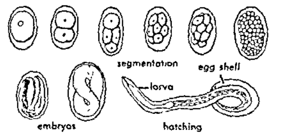
$$\frac{\text{Stress}}{\text{Strain}} = \frac{S}{\epsilon} = \text{constant} = E$$

where E is the modulus of elasticity, or Young's modulus, in pounds per square inch. The constant relationship between stress and strain applies only to stress below the proportional limit. For materials having a nonlinear stress-strain diagram, the law is an approximation applicable to low stress values. See STRESS AND STRAIN; YOUNG'S MODULUS.

[W.J.KR.]

Hookworm

Any of several parasitic roundworms of the class Nematoda, phylum Aschelminthes, which infect mammals. However, the name hookworm is usually used in referring to the Old World hookworm, *Ancylostoma duodenale*, or the New World hookworm, *Necator americanus*. Together these worms are the most serious parasites of man. They occupy most of the tropical and subtropical parts of the world. *Ancylostoma* is rare in the United States, but *Necator* is common in the southeastern states.



Development of the hookworm, Ancylostoma duodenale. (After Stiles from T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957)

Necator is somewhat similar to *Ascaris* in anatomy (see ASCARIS). Adults are from 8 to 15 mm long. Sexes are separate, the females being larger than the males. The inside of the mouth is equipped with several cutting plates by means of which the worm bites into the blood vessels lining the intestine. Blood is pumped from these wounds by the muscular pharynx. Each worm consumes about 0.8 milliliter of blood every 24 hours, along with some body fluids and bits of mucus from the intestinal wall. An anticoagulant enzyme is injected into the bite to keep the blood flowing freely. Frequently blood in excess of what the worm can digest is pumped from these bites.

Life history. In *Necator* the female produces about 9000 eggs per day over a period of 4 or 5 years. These pass to the ground with the feces of the host, where they hatch in 1-2 days. Temperatures of from 68 to 86°F and adequate moisture are required for the development of the embryo. The hatched larva undergoes two quick molts, feeding primarily on bacteria, and is then in the infective stage. They may enter by way of the mouth, passing directly to the intestine. Typically, however, they penetrate the skin, and are carried by the lymphatics and blood stream through the body to the lungs. Here they burrow through the lung membrane into the bronchioles, are coughed up, swallowed, and thus reach the intestine.

diseases may result from large infestations because of lung damage by the migrating larvae.

Occurrence. Hookworms are usually found in rural communities where there is poor sanitation. Until 1903 hookworm was a severe menace in the southern part of the United States. Vigorous preventive measures have now reduced it to a relatively uncommon disorder in most communities. Active

treatment of victims by vermifuges, the wearing of shoes, and proper sanitary practices are all important in hookworm control. In the tropics hookworm is still one of the greatest threats to the health of man. *See* HOOKWORM DISEASE.

There are a number of related species, including *Ancylostoma caninum* and *A. braziliense* which are almost cosmopolitan among cats and dogs in the warmer parts of the world. Both have been reported for man, but not commonly. Larvae of these species may cause "creeping eruption" in man when they are unable to reach the larger blood vessels and remain embedded in the skin *See* STRONGYLOIDEA. [J.D.B.]

Hookworm disease

Microcytic hypochromic anemia in man produced by the nematodes *Necator americanus* or *Ancylostoma duodenale* in the intestine. These nematodes, ⅜ in. long, each suck about 0.5 ml of blood daily from the small intestine. Although the clinical manifestations of both nematodes are similar, *A. duodenale* infections are usually more severe and less amenable to treatment. The host, if malnourished, cannot replace the blood lost and anemia ensues with all its symptoms.

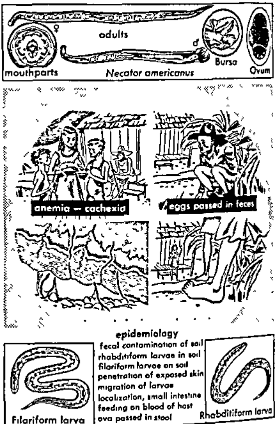
N. americanus lays about 10,000 eggs and *A. duodenale* up to 15,000 eggs daily. These eggs, which are identical, are eliminated in the feces. In poor sanitation areas, the soil becomes contaminated. With favorable temperature, humidity, shade, and aeration eggs in soil embryonate in 24 hours. A first-stage larva hatches and molts twice within a week to become infective. The larva actively penetrates the skin of man, gets into the blood stream, and then goes to the pulmonary alveoli. Penetration of the skin by infective hookworm larvae produces a dermatitis known as dew itch or ground itch. Migrating via trachea and esophagus to the duodenum, the parasite reaches adulthood in 2 months.

Diagnosis is made by examination of the feces. Eggs are counted to determine the severity of infection. Approximately 50 eggs per gram of feces represents a worm. Heavy infections show more than 10,000 eggs per gram.

Adult male farm workers are continuously exposed. Females and children are more apt to be exposed during harvesting. The disease was once a scourge in many tropical and subtropical lands, and was found frequently among miners and tunnel diggers in temperate climates. It has, however, decreased in importance with widespread control measures, improvement in sanitation and nutrition, and the wearing of shoes.

The drug tetrachlorethylene is used on an empty stomach, with or without purging, to remove the worms from the intestine.

The infective larva of *Ancylostoma braziliense*, found in dogs and cats, may infect the skin of man and produce a creeping eruption that becomes contaminated with bacteria. The condition is known as larva migrans. The lesion may subsist for



Epidemiology of hookworm disease. (*T. T. Mackie, G. W. Hunter, and C. B. Worth, A Manual of Tropical Medicine, 2d ed., Saunders, 1954*)

months. The indicated treatment involves freezing of the active end of the eruption with ethyl chloride and taking hetrazan, diethylcarbamazine citrate, by mouth. [J.F.MA]

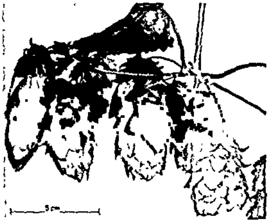
Hop

A dioecious liana, *Humulus lupulus*, belonging to the plant order Urticales. Each season numerous herbaceous vines are produced from a perennial crown. The vines twine in a clockwise direction, grow to a length of 15–25 ft in a season, and die following maturity. *See* STEM (BOTANY).

The inflorescences of the female plants constitute the hops of commerce. These inflorescences are catkins (strobiles) with papery bracts and brac- … granules … anules … value

in the brewing of beer and ale.

Hops are adapted to the temperate regions of the world. In the United States the major areas of production are Washington, Oregon, California, and Idaho. Hops are produced primarily under irrigation on fertile, well-drained soil. Rhizomes from the female plants are planted to produce 680–1030 hills/acre spaced 6½–8 ft apart. The vines are trained to climb strings suspended from an overhead wire trellis and attached to stakes at the hills (bases of plants). The hops are harvested



Hop female inflorescences. (USDA)

by specialized equipment or by hand, dried in large kilns, and baled for market. From 1948 to 1957 the average annual farm value of hops in the United States was $22,807,600.

Hops have been used by brewers for at least 1200 years. When added to the wort (fermenting material), they impart a characteristic flavor and aroma to the finished beverage and aid in preservation and protein coagulation. *See* MALT BEVERAGE; URTICALES. [S.N.B.]

Hophornbeam

The genus *Ostrya* of the birch family, represented in North America by two species. *O. virginiana*, a small tree which may reach a height of 60 ft, is widely distributed in the eastern half of the United States and in the highlands of southern Mexico and Guatemala. It can be recognized by its fruit which closely resembles that of the hop vine, and by its very scaly bark. The scales usually occur in narrow, more or less parallel, vertical strips. The winter buds are usually tinged with green showing about six striate scales, and the leaves are sharply and doubly serrate. This is one of several trees known as ironwood because of its hard, strong wood, and, like the hornbeam, it is used for fence posts, tool handles, mallets, and other articles requiring hardness and strength.

O. knowltonii is a small tree of rare occurrence, found in the southwestern United States. *See* FOREST AND FORESTRY; TREE. [A.H.G.]

Hoplocarida

The name of a superorder of the class Crustacea, with a single order, the Stomatopoda, commonly known as the mantis shrimps. The term Hoplocarida is used only in classification; the animal group as such is usually indicated with the name Stomatopoda *See* STOMATOPODA. [L.B.H.]

Hoplonemertini

An order of the class Enopla of the phylum Rhynchocoela. All species of Hoplonemertini have an armed proboscis which consists of an anterior thick-walled tube, a middle portion with stylets, and a posterior blind tube Two suborders are recognized, the Monostylifera with a single stylet and Polystylifera with many stylets. The middorsal vessel and lateral vessels are connected. Some common genera are *Emplectonema, Amphiporus, Carcinonemertes, Tetrastemma*, and *Geonemertes*. *See* ENOPLA. [C.B.C.]

Horizon

The visible horizon is the apparent boundary line between sky and earth or sea. The astronomical horizon is the great circle of the celestial sphere 90° from the zenith and the nadir. *See* ASTRONOMICAL COORDINATE SYSTEMS. [C.M.C.]

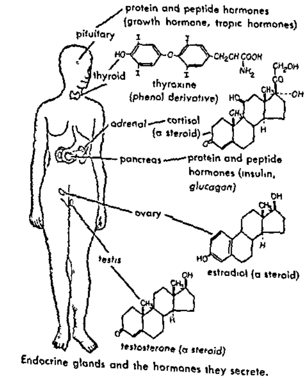
Hormone

One of the chemical messengers produced by endocrine glands, whose secretions are liberated directly into the blood stream and transported to a distant part or parts of the body, where they exert a specific effect for the benefit of the body as a whole. The endocrine glands involved in the maintenance of normal body conditions are pitui-

American hophornbeam, *Ostrya virginiana*. (A. H. Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

endocrine gland, elaborates a substance called secretin which stimulates the pancreas to produce its digestive juices. The placenta is also a very important hormone-producing tissue. See separate articles on the individual glands.

The hormones obtained from extracts of the endocrine glands may be classified into four groups according to chemical constitution: (1) phenol derivatives, such as epinephrine, norepinephrine, thyroxin, and triiodothyronine; (2) proteins, such as the anterior pituitary hormones, with the exception of ACTH, human chorionic gonadotropin, pregnant-mare-serum gonadotropin, and thyroglobulin; (3) peptides, such as insulin, glucagon, ACTH, vasopressin, oxytocin, and secretin; and (4) steroids, such as estrogens, androgens, pro-



Endocrine glands and the hormones they secrete.

gesterone, and corticoids. Hormones, all but a few exceptions like pituitary growth hormone and insulin, may also be classified as either tropic hormones or target-organ hormones. The former work indirectly through the organs or glands which they stimulate, whereas the latter exert a direct effect on peripheral tissues. Many of the target-organ hormones are steroids whose production and secretion are controlled by tropic hormones, which themselves are either proteins or peptides; for example, the target-organ steroid, cortisol, from the adrenal gland is released by the action of the tropic peptide hormone, pituitary ACTH. See PEPTIDE; PROTEIN; STEROID.

Hormones are responsible for the normal functioning of the body; thus, the organism's recurring needs call for a continuing cycle of hormonal production and circulation. The various hormones are secreted chiefly under the supervision of the anterior lobe of the pituitary gland, known as the master gland, and in terms of their physiological role, they are all either metabolic or gonad-stimulating; that is, they are involved either in the regulation of body chemistry or in the sex cycle. A key to the physiological importance of hormones is that these substances can manifest an effect even when present in very minute or trace quantities. For example, as little as 0.05 mg of the estrogenic hormone of the ovary will induce uterine bleeding in a woman.

This maintenance of the normal state of body with respect to the hormonal functioning of the various endocrine glands depends upon a feedback mechanism which tends to preserve an internal balance. The anterior pituitary gland and the adrenal cortex are parts of such a system. The pituitary hormone ACTH stimulates the adrenal cortex to secrete its steroid hormones. If one adrenal gland is removed, the ACTH stimulation causes the cortex of the remaining adrenal to increase in size. If, on the other hand, the anterior pituitary is removed, the consequent lack of ACTH results in considerable atrophy of the cortices of both adrenal glands. If ACTH is then injected, the adrenals regain their usual size. It appears that under normal conditions, the concentration of adrenal cortical hormones in the blood stream controls the secretion of ACTH by the pituitary; when this concentration is high, less ACTH is released; when it is low, the output of ACTH increases. The same sort of push-and-pull mechanism is believed to operate in the secretion of sex hormones by the ovaries and testes. See ALDOSTERONE; EPINEPHRINE; HORMONE, ADENOHYPOPHYSEAL; HORMONE, ADRENAL CORTEX; HORMONE, ADRENAL MEDULLARY; HORMONE, NEUROHYPOPHYSEAL; INSULIN. [C.H.L.]

Hormone, adenohypophyseal

Of all the endocrine glands, the anterior pituitary, or adenohypophysis, occupies the prime place because it controls the functions of the adrenal cortex, the thyroid, and the gonads. For this reason the hormones secreted by the anterior pituitary are called tropic hormones; that is, their function is the stimulation of a specific target organ, an endocrine gland, whose activity is thereby enhanced or awakened. This is true of all but one adenohypophyseal hormone, the growth hormone, which is secreted by the anterior pituitary, but which does not seem to have a specific target organ, or at any rate, a target organ has not been discovered. See HORMONE; PITUITARY GLAND.

There are six known anterior pituitary hormones whose existence has been firmly established. These six hormones can be classified into two groups, gonadotropic and metabolic. The gonadotropic hormones are follicle-stimulating hormone (FSH), interstitial-cell-stimulating hormone (ICSH), and lactogenic hormone (prolactin). The metabolic hormones are the thyrotropic hormone (thyroid stimulating hormone, TSH), adrenocorticotropic hormone (ACTH, adrenocorticotropin, corticotropin), and growth hormone (somatotropin, STH).

Gonadotropic hormones. In the female, FSH stimulates the development of immature ovarian follicles, bringing them to maturity. ICSH in turn stimulates the mature follicle to form estrogens, the estrus-producing hormones, and gives the follicle impetus for further development as a corpus luteum. The growing corpus luteum is then brought to maturity by lactogenic hormone, which, hence, is also known as luteotropic hormone (LTH). The lactogenic hormone at the same time stimulates the mammary gland, so that it is known by still another term, mammotropic hormone (MH).

The isolation of ICSH in apparently pure form from sheep and pig glands has been reported. Physicochemical and immunological studies indicate

that the hormones derived from the pituitary glands of these two species are not identical. Purified preparations of FSH have been obtained from sheep, pig, and human pituitaries. Of all the pituitary hormones, FSH appears to be the only one which is soluble in a half-saturated ammonium sulfate, $(NH_4)_2SO_4$, solution. Both FSH and ICSH preparations appear to contain carbohydrate, that is, hexose and hexosamine, in addition to constituent amino acids. *See* AMINO ACIDS; CARBOHYDRATE.

Lactogenic hormone has been isolated in highly purified form from both sheep and beef pituitary glands; the sheep prolactin can be differentiated from the beef hormone by differences in solubility behavior and in tyrosine content. Both hormones have molecular weights of 26,000 and isoelectric points at pH 5.7; each consists of a single peptide chain with 211 amino acid residues. It has further been shown that the prolactin molecules have the sequence threonylprolinylvalinylthreonylproline (Thr.Pro.Val.Thr.Pro.) at the nitrogen or N terminus, and an intrachain disulfide loop at the carbon or C terminus.

Metabolic hormones. The anterior pituitary hormones which control and regulate body metabolism are the thyrotropic, adrenocorticotropic, and growth hormones. The thyrotropic hormone stimulates the thyroid gland to secrete its hormone, thyroxin, and ACTH stimulates the adrenal cortex to produce cortisol and other adrenocortical hormones. The growth hormone influences the general process of body growth. The metabolic functioning of the body depends on the balance or the interaction among these hormones.

Purified TSH preparations have been obtained from beef pituitary glands. Observations that the TSH activity is soluble in the presence of certain protein-precipitating agents would seem to indicate that the hormone may possess a low molecular weight. There is also evidence to indicate that TSH may be a glycoprotein.

From clinical observations of disturbances in the growth process, which are possibly the most obvious to perceive of all endocrine results, it was evident that the pituitary gland bears a direct relationship to body growth. Early specific experiments confirmed these general clinical observations, offering a convincing argument that the anterior pituitary secretes a hormone which possesses growth-promoting activity. Final proof came with the isolation, in highly purified crystalline form from beef pituitary glands, of a hormone known, because of its obvious effects on growth, as growth hormone, or somatotropin.

Although the effects of the bovine growth hormone were readily demonstrable in the rat and the mouse, it seemed to be ineffective in humans. Evidence gathered by a number of investigators pointed to an important connection between the biological effectiveness of the hormone and the species from which it was derived. For example, it was found that the bovine hormone is inactive not only in humans but also in monkeys, whereas the hormone preparation from monkey glands is effective in that particular species. In the same way, growth hormone concentrate prepared from fish pituitaries is effective in fish, but not in rats. The bovine hormone, however, is effective in both fish and rats, although it has also been reported to have no effect in the guinea pig. All this evidence suggested that the growth hormones isolated from various species may not be chemically identical. This was confirmed in 1956 when highly purified preparations of growth hormone were successfully isolated from both human and monkey pituitaries. Chemical studies on these preparations demonstrated a difference between the primate growth hormones and the bovine hormone; for example, the former hormones have a lower molecular weight and a more acidic isoelectric point. Furthermore, structural investigations show that the bovine growth hormone molecule has the form of a branched polypeptide chain, whereas the human and monkey hormones consist of a straight polypeptide chain. A growth hormone prepared from whale pituitaries has also been reported and has been shown to have different physical and chemical characteristics from the growth hormones of other species. Structurally, the whale growth hormone is closer to the human than it is to the bovine hormone. The table summarizes the physicochemical properties of highly purified growth hormones from different species. In this table, Phe is phenylalanine; Thr, threonine; Ala, alanine; Ser, serine; Leu, leucine; and Gly, glycine.

In 1957, monkey and human growth hormones prepared by different methods in various laboratories were administered to human subjects, including hypophysectomized and dwarf patients. In these clinical trials, the effectiveness of the primate hormones in human subjects was clearly demonstrated, in contrast to earlier clinical failures with the bovine hormone. It has also been demonstrated that a single human pituitary gland, whose average weight in adult humans is about 750 mg, contains approximately 3–5 mg of growth hormone.

The removal of the anterior pituitary in almost all the species studied brings about a pronounced atrophy, or reduction in size, of the cortex of the adrenal glands, whereas the medulla is affected scarcely at all. On the other hand, implantation of anterior pituitary tissue or injection of pituitary extracts into normal animals has, as its chief effect, a marked hypertrophy of the adrenal cortex. The name adrenocorticotropic hormone (ACTH) or adrenocorticotropin has been employed to designate the active principle in pituitary extracts that accomplishes repair of adrenal-cortical atrophy following hypophysectomy, removal of the pituitary gland. In both experimental animals and [illegible] of ACTH accelerate the [illegible]

hormone has been employed clinically for remission of rheumatoid arthritis and other meta-

Some characteristics of various pituitary growth hormones

Properties	Beef	Sheep	Whale	Monkey	Human
Molecular weight	45,000	48,000	40,000	25,000	27,000
Isoelectric point, pH	6.8	6.8	6.2	5.5	4.9
No. of —S—S— bridges	4	5	3	4	2
N-terminal sequence	Phe.Thr.Ala.... Ala.Phe.Ala....	Phe.... Ala....	Phe....	Phe.....	Phe.Ser.Thr ..
C-terminal sequence	...Leu.Ala.Phe.Phe.	.. .Ala.Leu.Phe.	....Leu.Ala.Phe	....(Ala,Gly).Phe	....Leu Phe

bolic disorders, as well as for a diagnostic agent in tests of adrenal function.

ACTH has been isolated in pure form from pig, sheep, and beef pituitaries; although all three ACTH preparations possess similar biological properties, there are slight differences among them in amino acid composition and in chemical structure, as summarized below:

Structure of bovine ACTH:

Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Try-Gly-Lys-Pro-
1 2 3 4 5 6 7 8 9 10 11 12

Val-Gly-Lys-Lys-Arg-Arg-Pro-Val-Lys-Val-Tyr-Pro-
13 14 15 16 17 18 19 20 21 22 23 24

Asp-Gly-Glu-Ala-Glu-Asp-Ser-Ala-Glu(NH_2)-Ala-Phe-Pro-
25 26 27 28 29 30 31 32 33 34 35 36

Leu-Glu-Phe
37 38 39

Species	Amino acid residue on carbon atom number 25 26 27 28 29 30 31 32 33
Pig	Asp-Gly-Ala-Glu-Asp-Glu(NH_2)-Leu-Ala-Glu
Sheep	Ala-Gly-Glu-Asp-Asp-Glu-Ala-Ser-Glu(NH_2)
Beef	Asp-Gly-Glu-Ala-Glu-Asp-Ser-Ala-Glu(NH_2)

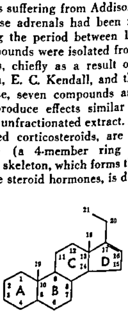
Ser is serine; Tyr, tyrosine; Met, methionine; Glu, glutamic acid; His, histidine; Phe, phenylalanine; Arg, arginine; Try, tryptophan; Gly, glycine; Lys, lysine; Pro, proline; Val, valine, Asp, aspartic acid; Ala, alanine; and Leu, leucine.

The adrenocorticotropins from these species are single-chain polypeptides composed of 39 amino acids with serine and phenylalanine as N- and C-terminal residues, respectively. The only difference in amino acid composition between the pig and sheep hormones appears to be one more leucine in the former and one more serine in the latter; this difference occurs in the sequence alanine-serine in carbon atom positions 31 and 32, which appears in the sheep peptide in place of the leucine and alanine which appear in the same positions in the pig hormone. Although there are no differences in amino acid composition between the sheep and beef hormones, a difference in amino acid sequence that occurs in a small portion of the polypeptide chain seems to support the conclusion that sheep and beef adrenocorticotropins are distinct chemical entities. [C.H L.]

Hormone, adrenal cortex

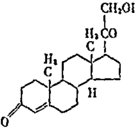
The cortex of the adrenal glands is essential to life. Absence of adrenal cortex hormones following destruction of the glands by tubercular infection or bilateral adrenalectomy, was shown by T. Addison (1855) and C. E. Brown-Sequard (1856), respectively, to lead to marked disturbances in electrolyte, carbohydrate, protein, and fat metabolism, and eventually to result in death. Maintenance of the structure, growth, and secretory activity of the adrenal cortex is under the control of the adrenocorticotropic hormone (ACTH) which is a polypeptide hormone of the anterior pituitary.

In 1930, W. W. Swingle, F. A. Hartman, and their coworkers described methods of preparing adrenal cortical extracts that were capable of keeping alive patients suffering from Addison's disease and animals whose adrenals had been removed surgically. During the period between 1935 and 1957, over 40 compounds were isolated from adrenal cortical extracts, chiefly as a result of the work of T. Reichstein, E. C. Kendall, and their collaborators. Of these, seven compounds are biologically active and produce effects similar to those produced by the unfractionated extract. All these compounds, called corticosteroids, are derivatives of phenanthrene (a 4-member ring hydrocarbon) whose carbon skeleton, which forms the basic structure of all the steroid hormones, is depicted as follows:

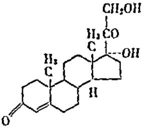


Phenanthrene

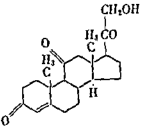
Corticosterone and cortisol are the major corticosteroids secreted by the adrenal cortex into the circulation; deoxycorticosterone, 17-hydroxydeoxycorticosterone, 11-dehydrocorticosterone, cortisone, and aldosterone, the other corticosteroids, are secreted in much smaller amounts. The formulas of these corticosteroids, with the exception of aldosterone, are as follows:



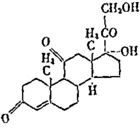
Deoxycorticosterone



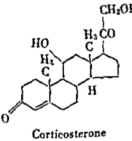
17-Hydroxydeoxycorticosterone



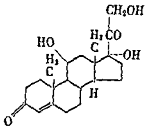
11-Dehydrocorticosterone



Cortisone



Corticosterone



Cortisol

The hydroxy (OH) group on the carbon atom 11 of the molecule in these biologically active corticosteroids rises above the plane of the ring structure. This is known as the β position and is shown as a solid line, whereas the 17-hydroxyl group in the α position, that is, below the plane of the ring system, is designated by a broken line in the formula.

The dramatic discovery by P. S. Hench and his associates in 1948 of the beneficial effect of cortisone on the inflammatory and allergic reactions in rheumatoid arthritis and other collagen diseases demonstrated that the hormones of the adrenal cortex exert a profound effect not only on intermediary metabolism, but on host resistance and reaction to disease as well. Cortisone, 17-hydroxy-11-dehydrocorticosterone, was first isolated independently by E. C. Kendall and by T. Reichstein, O. Wintersteiner, and their collaborators in 1936. The total synthesis of the hormone was achieved by R. B. Woodward in 1951. Numerous partial and total syntheses have since been effected by chemical and microbiological methods.

For bioassay of the active corticosteroids, the following procedures, employing rats as experimental animals, are generally used: (1) decrease in the ratio of sodium to potassium ions in the urine of adrenalectomized rats, (2) promotion of the survival and growth of young adrenalectomized rats, and (3) increase in the liver glycogen of fasted rats. The approximate relative biological potency of the seven corticosteroids as determined by the above bioassay methods is given in the table, together with their melting points (mp) in degrees centigrade and their specific optical rotation ($[\alpha]_D^0$). In column 2, the names of these corticosteroids given by Reichstein (R) and Kendall (K) are also listed.

excess, they increase gluconeogenesis, which is the formation of glucose and glycogen from nonglucose matter. This, in turn, may cause a diabetic state. In addition, they exert a moderate antiinsulin

Relative biological potency of corticosteroids

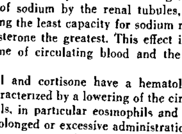
Corticosteroids	Substance		Formula	MP, °C	$[\alpha]^0_D$	Relative activity		
	R	K				Urinary Na^+/K^+	Survival and growth	Glycogen deposition
Cortisol	M	F	$C_{21}H_{30}O_5$	220	+167	0.8	0.5	16
Cortisone	F	E	$C_{21}H_{28}O_5$	215	+209	0.6	2.5	10
Corticosterone	H	B	$C_{21}H_{30}O_4$	182	+223	1.4	1.7	5
11-Dehydrocorticosterone		A	$C_{21}H_{28}O_4$	180	+229		1.0	5
17-Hydroxydeoxycorticosterone	S		$C_{21}H_{30}O_4$	205	+132	0.8		
Deoxycorticosterone	Q		$C_{21}H_{30}O_3$	142	+178	10	10	0.1
Aldosterone						1000		3

effect by blocking the action of the latter hormone at the periphery of the cell All biologically active corticosteroids are capable of increasing the reabsorption of sodium by the renal tubules, cortisol manifesting the least capacity for sodium retention and aldosterone the greatest. This effect increases the volume of circulating blood and the urinary output.

Cortisol and cortisone have a hematologic activity characterized by a lowering of the circulating blood cells, in particular eosinophils and lymphocytes. Prolonged or excessive administration of the hormone may cause polycythemia, a disease due to the excessive production of red corpuscles in the blood. Cortisol and cortisone are known to increase the secretion of hydrochloric acid by the stomach. An increased excitability of the central nervous system following the administration of cortisol or cortisone has also been reported. [C.H.L.]

Hormone, adrenal medullary

The adrenal medulla secretes two hormones: epinephrine or adrenalin and norepinephrine or noradrenalin. Both of these adrenal medullary hormones are catechol amines; epinephrine differs from norepinephrine only by the addition of an *N*-methyl group:



Epinephrine Norepinephrine

The discovery by L. Oliver and H. E. Schafer in 1894 that the blood pressure of an experimental animal was markedly increased by an intravenous injection of adrenal medullary extracts led to the isolation in pure form of epinephrine in 1901 by J. Takamine. Epinephrine is a colorless crystalline compound which is optically active; *that is, it causes a plane of polarized light to rotate, indicating centers* of asymmetry in the molecule. It is comparatively stable in acid solution, but in the presence of alkali, it combines readily with oxygen. The synthesis of epinephrine was first achieved by F. Stolz in 1904.

Although norepinephrine was shown by U. S. von Euler in 1946 to be secreted at the sympathetic nerve endings and to serve as sympathetic nerve transmitter, it was not until 1949 that S. Bergstrom isolated the hormone in the form of a crystalline base from the adrenals of cattle, where it occurs together with epinephrine in the approximate proportions of 1:4. In humans, the adrenal medulla contains 2.4 to 4 milligrams (mg) per gram (g) of both these hormones together; 10–30% of this amount is norepinephrine. In other species, for example, cat, toad, and pigeon, the amounts of epinephrine and norepinephrine in the adrenal glands are nearly equal. Small quantities of epinephrine may be found in nerve extracts, but there is no evidence that it is produced by the adrenergic or sympathetic nerve fibers.

Both epinephrine and norepinephrine raise the blood pressure, but by different mechanisms. Epinephrine causes an increase of cardiac output and myocardial irritability, and generalized vasodilatation. On the other hand, norepinephrine causes marked peripheral vasoconstriction with no alteration of cardiac output or myocardial irritability. Epinephrine can produce pronounced excitation of the organism by stimulation of the nervous system, whereas norepinephrine given in similar doses does not. Levels of circulating eosinophils, which are cells in the blood readily stained by eosin, are acutely depressed by epinephrine, but only slightly by norepinephrine. Epinephrine also exercises metabolic effects on oxygen consumption, basal metabolic rate, and hepatic glycogenolysis that are much greater than those of norepinephrine. The striking differences in the activities of the two adrenal medullary hormones are summarized below.

Epinephrine	*Norepinephrine*
Vasodilatation	Vasoconstriction
Marked action on metabolism	Metabolic effect only 1/4 to 1/10 that of epinephrine
Pronounced central excitation	Weak central excitation
Multifarious action and emergency function	Regulation of normal blood pressure

See ADRENAL GLAND; EPINEPHRINE; HORMONE. [C.H.L.]

Hormone, neurohypophyseal

Extracts of the neurohypophysis or posterior lobe of the pituitary possess the following biological properties: (1) effect on blood pressure by constriction of the blood vessels leading from the arteries, known as pressor action, (2) stimulation of smooth muscles and specifically of the musculature of the uterus, known as oxytocic action, and (3) inhibition of urine excretion through the kidneys, known as antidiuretic action. These effects of neurohypophyseal extracts are due to the presence of two peptide hormones, oxytocin and vasopressin.

Two biologically active principles were separated by O. Kamm and his colleagues in 1928, and have been available commercially. The vasopressor principle has been called pitressin and the uterine-stimulating principle has been called pitocin. The commercial unfractionated extract of the posterior lobe has been called pituitrin. For some time, however, it was not clear whether the three activities were the result of two or three distinct chemical entities. It was not until the isolation, structure, and synthesis of the two hormones by V. du Vigneaud and coworkers between 1950 and 1954 that it was clear that vasopressin is responsible for both the pressor and antidiuretic effects, whereas uterine stimulation is the principal effect characteristic of oxytocin. However, a study of these two neurohypophyseal hormones, assayed according to various methods, by H. B. van Dyke and his colleagues, has demonstrated that in fact there is overlapping of the biological activities of the two. These data have been summarized in the table below:

Potency of purified oxytocin and vasopressin in U.S. Pharmacopeia units per milligram*

	Oxytocic (rat uterus)	Avian depressor (fowl)	Milk ejecting (rabbit)	Pressor (rat)	Anti-diuretic (dog)
Oxytocin	500	500	500	7	3
Vasopressin	30	85	100	600	600

*From G. Pincus (ed.), *Recent Progress in Hormone Research*, 11:4, 1955.

From a quantitative standpoint, vasopressin is the more versatile of the two, since it exercises appreciable uterine-stimulating, avian vasodepressor, and milk-ejecting actions in addition to its pressor and antidiuretic effects. Oxytocin has a slight but definite pressor and antidiuretic action in addition to its own characteristic effects. These overlapping biological activities of the two hormones reflect their close relationship in chemical structure.

Both hormones are amides; that is, they are organic compounds in which an acidic or carboxyl group is combined with ammonia. The hormones are composed of a cystine-containing pentapeptide in which a portion of a molecule is made up of five amino acids linked to a tripeptide amide side chain, which contains three amino acids. Their structural formulas are depicted as follows:

CyS-Tyr-Ileu-Glu(NH$_2$)-Asp(NH$_2$)CyS-Pro-
Leu-Gly(NH$_2$)

Oxytocin (ox and hog)

CyS-Tyr-Phe-Glu(NH$_2$)-Asp(NH$_2$)-CyS-Pro-
Arg-Gly(NH$_2$)

Vasopressin (ox)

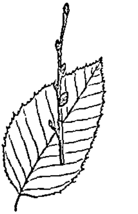
where CyS is cysteine; Tyr, tyrosine; Ileu, isoleucine; Glu, glutamic acid, Asp, aspartic acid; Pro, proline; Leu, leucine; Gly, glycine; Phe, phenylalanine, Arg, arginine.

In vasopressin obtained from hog glands, a lysine residue is located in the position of the arginine residue of the bovine hormone However, oxytocin preparations from either species have the same structure. The isoelectric point of oxytocin is located at pH 7.7, and that of bovine vasopressin at about 10.9. The molecular weight of oxytocin is 1007, and that of bovine vasopressin is 1084. *See* AMINO ACIDS, HORMONE; PITUITARY GLAND.

[C.H.L.]

Hornbeam

The genus *Carpinus* of the birch family, represented in the United States by *C. caroliniana*, the American hornbeam or blue beech. Hornbeam is a small tree sometimes attaining a height of 35 ft and has a smooth, steel-gray, fluted bark. It grows throughout the eastern half of the United States, especially in moist soil along banks of streams; it is sometimes called water beech When mature, it is easily recognized by its peculiar bark, by the doubly serrate leaves resembling those of sweet birch, and by the small, pointed, angular winter ... The fruit is a small ... ract. The ... is some- ... is tree is rarely used for sawn products, being short and



American hornbeam, *Carpinus caroliniana*. (A. H. Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

usually crooked. The leaves turn scarlet or orange in the fall.

The European hornbeam, *C. betulus*, is often cultivated in parks and estates. It can be distinguished by its larger size, larger winter buds, and larger 3-lobed, almost entire fruiting bracts. *See* FOREST AND FORESTRY; TREE. [A.H.C.]

Hornblende

A general name given to the monoclinic calcium amphiboles that form extensive solid-solution series between the various metals in the generalized formula $(Ca,Na)_2(Mg,Fe,Al)_5(Al,Si)_8O_{22}(OH,F)_2$. Hornblende has a widespread occurrence in metamorphic, igneous (intrusive), and volcanic rocks ranging from dominantly ferromagnesian rocks to granites. The term common hornblende refers to the middle metamorphic grade hornblendes found in schists. Common hornblende usually forms as long prismatic needles, dark green to black in color. The hornblendes of the high grades of metamorphism are usually short, stubby prisms, black to brown in color and high in iron, aluminum, and some sodium. *See* AMPHIBOLE; METAMORPHISM.

Hornblende exhibits the characteristic 56° amphibole (110) cleavages and, except for tremolite, is green, black, or brown. In thin sections, hornblende is strongly colored (greens and brown) and strongly pleochroic (color change on rotation in plane-polarized light); the sodium hornblendes, when low in iron, are blue or bluish-green. Common hornblende is associated with plagioclase, garnet, epidote, pyroxenes, quartz, chlorite, anthophyllite and biotite; it also forms monomineralic masses. The aluminum- and iron-rich hornblendes are more often found with the alkali feldspars, quartz and biotite, in granites, gneisses, and pegmatites. The sodium-rich varieties (except glaucophane) are found in the alkali-rich rocks. A textural variety of intertwinned needles near tremolite in composition is called nephrite and is used in place of jade for carvings. *See* GLAUCOPHANE; TREMOLITE; *see also* JADE. [G.W.D.]

Horned toad

Any of several short, flat lizards of the genus *Phrynosoma*, family Iguanidae. The horned toads have short tails and are covered with rough scales, many of which are enlarged into spines. The head of most species is margined with spines, giving these lizards an appearance wholly unlike that of any other known animal except a similar-appearing but unrelated genus in Australia.

Horned toads are best known for their ability to squirt blood from their eyes when excited, a well-known trait in several species. They cannot, however, live for years sealed in airtight containers, as popular superstition supposes. Horned toads may be either oviparous or ovoviviparous, sometimes varying in this habit between individuals of the same species. There are seven species in the southwestern United States, and several others in Mexico. *See* SQUAMATA. [J.D.B.]

Hornet

A large species of wasp of the family Vespidae, order Hymenoptera, *Vespa maculata*, often called the bald-faced hornet. This insect is a paper wasp, closely related in its structure and habits to the other paper wasps and yellow jackets. It is distinguished by its large size, its black body marked with white and yellow markings, and the nature and size of its nest. It occurs throughout North America, but is somewhat more common in the South.

White-faced hornet, *Vespa maculata*, approx. 1 in long. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

The hornet nest is composed of several tiers of paper nests, each layer similar to the nest of the common paper wasps, but enclosed in a thick covering of paper combining the whole structure into an ovoid mass up to 12 inches in diameter and 18 to 20 inches long. This nest is suspended from a tree branch or frequently from the eave or side of a house. The principal opening is at the bottom. Hornets are predaceous upon other insects, and feed their larvae chewed up insects.

Hornets are peaceful if undisturbed, but will attack fiercely in a group if their nest is disturbed. They inflict a very painful sting. As many as 15,000 insects may live in a single nest. *See* HYMENOPTERA; INSECTA; WASP. [J.D.B.]

Hornfels

A common name given to a class of metamorphic rocks produced by contact metamorphism, also known as hornstone. Hornfelses were originally sedimentary rocks. As magma intruded into the sediments the heat given off induced recrystallization and effected a complete alteration of the primary sedimentary strata into hard, often flinty rocks. Their fine-grained constituent minerals usually can be discerned only with the microscope. Chemical alterations accompanying the formation of normal hornfelses are small, except for the partial removal of certain fugitive constituents of the sediments (such as water and carbon dioxide). Thus, the chemical composition of a hornfels depends only upon that of the original sediment, and consequently the mineral composition of a hornfels is predetermined by the nature of the original sediment. *See* METAMORPHISM; METAMORPHIC ROCKS.

Among the varieties of hornfelses which may develop from various sediments the continuous series from shale to limestone with some admixture of marl is the most interesting. Chemically this series is made up of the following chief constituents: silicon dioxide, SiO_2, aluminum oxide, Al_2O_3, calcium oxide, CaO, iron(II) oxide, FeO, and magnesium oxide, MgO. Assume first that there is sufficient SiO_2 to form highly silicified minerals and secondly that FeO and MgO can be grouped together because they substitute for each other diadochally. Other than silica this is a system of three variables, alumina, lime, and ferromagnesia, as illustrated in Fig. 1. According to the chemical composition of the sediments there will be 10 classes of silica rich hornfelses:

1. Pure shale contains alumina and traces of ferromagnesia which form andalusite and cordierite (and an excess of silica which produces quartz).
2. A small addition of lime produces anorthite (in the actual hornfels represented by plagioclase, since some Na_2O is present) in addition to andalusite and cordierite.
3. More lime produces more anorthite, whereas andalusite disappears.
4. Next hypersthene will form in addition to cordierite and anorthite.
5. With still more lime cordierite disappears, with formation of more anorthite and hypersthene.
6. Lime is now so dominating that it combines with magnesia to form diopside in addition to anorthite and hypersthene.
7. Hypersthene disappears.
8. More lime reacts with anorthite to form grossularite in addition to diopside.
9 Anorthite is completely used up in the grossularite reaction.

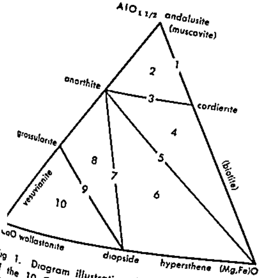


Fig 1. Diagram illustrating the mineral assemblages of the 10 Goldschmidt classes of hornfelses. Other mineral constituents, in addition to those shown in the diagram, are quartz, orthoclase, and accessories such as apatite

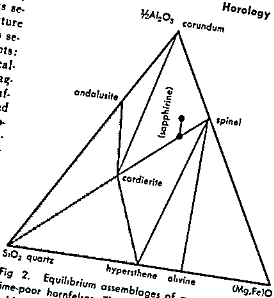


Fig 2. Equilibrium assemblages of mineral phases in lime-poor hornfelses. The stippled area represents the field of normal pelitic sediments. (After C. E. Tilley, *Mineral. Mag.*, 1923)

10 Pure lime silicate, that is, wollastonite, occurs with diopside and grossularite.

In certain regions, however, limestone is missing as a component in the sediments, and the corresponding hornfelses will be poor in lime, with a composition approaching the ternary system SiO_2-Al_2O_3-(Mg,Fe)O. Figure 2 shows the mineral associations encountered in such hornfelses. Each of the seven triangular areas shown in Fig. 2 corresponds to a possible mineral assemblage. Two of the assemblages, quartz-cordierite-andalusite and quartz-cordierite-hypersthene, are representatives of the usual silica-rich pelitic hornfelses. The other mineral assemblages occur in hornfelses of the Hartz Mountains, Germany, of the Cambric area, Perthshire, Scotland, and elsewhere.

The natural hornfelses very closely follow the theoretical scheme of mineral associations, showing incontestably that a state of chemical equilibrium was approached in the rocks and that the laws of physical chemistry may be applied to metamorphic processes.

[T.F.W.B.]

Horology

The science of measuring time and the technology of constructing instruments for the measurement of the passage of time. Watch and clock making is also termed horography. Early timepieces used the shadow cast by the sun (*see* SUNDIAL). They were fixed on stands or to the sides of buildings; others were small and portable, some with adjustments for days of the year and latitude of the user. Other forms of timepiece included the water clock, sand (hour) glasses, and candles graduated in hours.

Mechanical clocks date from the fourteenth century. In them the fall of a weight provides motive power to overcome friction in the time train of

gears and to actuate a pendulum. The pendulum, in turn, regulates the fall of the weight (*see* PENDULUM). Later, the spring was invented and used as the motive power, and the balance wheel escapement replaced the pendulum (*see* WATCH). The timepiece may be highly refined and supported to isolate shock and vibration. *See* CHRONOMETER.

Electrical clocks, driven by synchronous motors, rely upon the steadiness of the electric generator for their accuracy (*see* RELUCTANCE MOTOR). The generator speed is usually governed mechanically at the central station (*see* ELECTRIC POWER GENERATION). Crystal clocks use the mechanical resonance of a crystal plate coupled piezoelectrically into an electronic circuit (*see* PIEZOELECTRIC CRYSTAL). Atomic, or, more properly, molecular clocks are regulated by the natural resonance produced by the mass of an atom and the intramolecular force upon it in a molecule, the energy of the action in vibration being coupled electromagnetically into the circuit of the clock. *See* ATOMIC CLOCK.

Traditionally, time measurement is based on the rotation of the earth on its axis (*see* ASTRONOMICAL INSTRUMENTS). However, as electronic methods become more precise, it is possible to observe slight fluctuations in the movement of the earth. *See* EARTH (ORBITAL MOTION); TIME. [F.H.R.]

Horse

A term commonly applied to the domestic horse, *Equus caballus*, but sometimes used in a broader sense to include the hoofed mammals of the family Equidae, represented throughout the world by many fossil and living species, such as the donkey, zebra and quagga. *See* DONKEY; ZEBRA.

The domestic horse is now developed into a wide array of breeds, including several small forms which are commonly called ponies. All domestic horses are thought to have been derived from Przewalski's wild horse, *Equus przewalskii*, still living as a wild animal in Mongolia, and improved through crossing with the now extinct Libyan horse, *E. libycus*, and possibly others. *See* HORSE PRODUCTION; MULE; PERISSODACTYLA. [J.D.B.]

Horse production

A specialized phase of animal agriculture in the United States. Although formerly used by millions of farmers and by the U.S. Army as a source of power, horses are now used by Americans principally for sport and pleasure. Tractors and modern farm machinery have replaced work horses on most farms. However, the breeds of light horses still flourish; horse racing is one of the leading spectator sports. Horse shows throughout the United States and horsemanship training at many schools and colleges attest to the popularity of the light horse.

Horse production is a major enterprise on a few farms; a sideline business or a hobby on many others. The bluegrass region of central Kentucky is the most famous American horse-breeding region where a large proportion of the best Thoroughbred studs, horse-breeding farms, and outstanding breeders of Standardbred and American saddle horses is concentrated. However, horses are produced in all the states and in greatest numbers in Texas. Cattle ranches of the West and Southwest use many horses.

Successful horse production depends upon careful selection and mating of breeding stock, proper feeding, careful management including parasite and disease control, and profitable selling of the horses produced in the breeding program.

Horse-breeding practices. These vary with the type and value of the animals involved. However, all horsemen select breeding stock on the bases of individuality, performance, pedigree, and progeny. Young animals are chosen for breeding purposes chiefly on performance in a particular field of capability. Preference is shown for young horses from bloodlines that have consistently produced the desired type and ability. After his horses have reproduced, the good horse breeder retains or culls them on the basis of the quality of their progeny.

Domesticated horses can mate at any season of the year, but for practical purposes the horse-breeding season in the United States is limited to late winter, spring, and early summer. Breeders prefer to have their mares foal in the spring when the pasture season begins. The pregnancy period lasts roughly 11 months. Therefore, mares bred in the spring months produce their foals the following spring.

Although yearling stallions are capable of breeding mares, good horsemen never use a stallion for breeding service until he is at least 2 years old. Many colts (young stallions) when 2 and 3 years of age are still being trained under saddle or in harness. Therefore, in practice most stallions of the light breeds begin stud service at 4.

Breeders of Thoroughbreds in the United States usually limit their stallions to about 40 mares a year. In his first breeding season, a colt is bred to only 15 or 20 mares. Some Standardbreds and stallions of other breeds may breed 100 mares or more over a long breeding season. Breeding service is permitted once or twice a day, rarely more often. Records kept at many stud farms show that 2–3 services are required for every pregnancy. A well-managed stallion bred to mares that are normal in every way should get about 80% of his mares with foal.

Good feed and care are necessary to maintain a high degree of fertility in the stallion. During the breeding season a stallion requires about 1 lb of grain daily for every 100 lb of his body weight. Oats are the principal grain, supplemented often with wheat bran and linseed meal (*see* FLAX; OATS; WHEAT). In addition to the grain, a stallion needs 1¼ lb of hay for every 100 lb of body weight (*see* GRASS CROPS). At least one-half of the hay should be a good quality legume, such as alfalfa (*see* ALFALFA; LEGUME). Green forage helps to promote high fertility. Therefore, a stallion should have opportunity to graze in a paddock or pasture. The

Fig. 1. (a) Percheron mare; (b) Belgian stallion (J. F. Abernathy Live Stock Photo. Co.). (c) Clydesdale stallion; (d) Shire stallion; (e) Suffolk stallion (USDA photographs).

most successful horsemen exercise their breeding stallions by riding or driving them daily during the breeding season.

Although in the western part of the United States a horseman may allow his stallion to run with a band of mares, in the rest of the country hand breeding is the usual practice. The stallion and mare are bridled, and the mating is controlled by the men in charge.

The volume of semen produced by a stallion at time of mating ranges from 30 to 150 ml. The average number of sperm cells is 60,000 per cubic millimeter.

Because only one stallion is needed for a large number of mares, many stallions that are not desired for breeding are castrated. Following castration, they are known as geldings. Colts are usually castrated when they are yearlings. Geldings are often more docile and more easily handled than stallions.

Young fillies reach puberty at about 1 year of age. However, they are not bred until they are 3 or 4 years old, or later, if they are being raced or shown. Mares have an estrous cycle of 21 days. The period of estrus, or heat, may occupy 3–7 days of this cycle. Ovulation occurs approximately 2 days before the end of the heat period. Horsemen try to mate their mares close to the time of ovulation. Where horses of value are being bred, all possible sanitary measures are followed at the time of mating. If a mare conceives, she usually does not come into heat during her pregnancy.

Artificial insemination of mares is not widely practiced in the United States. *See* BREEDING (ANIMAL). Many of the horse-breeding associations do not permit registration of foals produced by artificial breeding. Furthermore, the long heat period of the mare makes the optimum timing of artificial insemination difficult.

The average gestation period for mares is 340 days. For the United States, only about 55% of the mares that are bred produce live foals. However, under ideal conditions on some Kentucky stud farms, 70–80% of the mares produce foals.

A normal foaling may take only 10–20 minutes. Newborn foals weigh from 6 to 10% of the weight of their dams.

Types of horses bred. In the United States these include work horses, heavy harness horses, light harness horses, and saddle horses.

Five breeds of work horses were once used. Only two, however, were numerous. These were the Percheron and the Belgian (Fig. 1). Percherons, originally from France, are black or gray, while Belgians are chestnut or roan. Clydesdales from Scotland, and Shires and Suffolks from England were the other breeds, but were never widely popular in this country. The mechanization of American agriculture has rendered the work-horse obsolete in many farming areas.

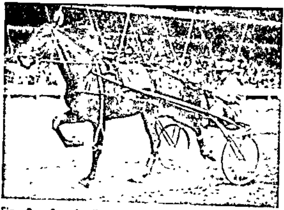


Fig. 2. Standardbred trotter (*Photograph by The Harness Horse*)

Heavy harness horses were the coach horses or carriage horses of old. The chief representative of this type in the United States is the Hackney, an English breed noted for its high-stepping trot. Hackneys are now seen only in some of the larger American and Canadian horse shows.

Light harness horses are represented in the United States by an American breed, the Standardbred (Fig. 2). These horses are numerous and popular as harness race horses. Some Standardbreds race at the trot, a two-beat gait in which the diagonal legs work together. Others race at the pace, a two-beat gait, where both left legs and then both right legs move in unison. Most harness races are for a distance of 1 mile, and the horse pulls a light, two-wheeled sulky.

Many breeds used under saddle are popular in the United States. The oldest is the Thoroughbred, developed in England in the eighteenth century (Fig. 3). Thoroughbreds are used for racing on the flat, for steeplechasing, for polo, and for fox hunting.

The American quarter horse, a breed developed in the Southwest, is the only American horse with work to do outside of sport (Fig. 4). This breed is used on the cattle ranches of the West and Southwest. Quarter horses are also used for short racing. The breed name comes from the fact that these horses are very fast up to a quarter of a mile, but do not maintain their speed over a long distance.

Fig. 3. Thoroughbred stallion. (*Photograph by The Thoroughbred Record*)

The American saddle horse breed was developed in Kentucky and Missouri. The horses are used for pleasure riding. Some are exhibited as three-gaited saddle horses and perform the walk, trot, and canter (slow gallop). Others taught two additional gaits, the rack and slow gait, are shown as five-gaited saddle horses (Fig. 5). Tennessee walking horses, Morgans, Arabians, Palominos, and Appaloosas are other saddle horses ridden for pleasure in various sections of the United States.

During the 1950s pony breeding in the United States increased tremendously (Fig. 6). Shetlands are the most popular. Purebred Shetland ponies registered in 1958 outnumbered horses of all breeds except Thoroughbreds and American quarter horses. Welsh ponies, also, have increased in favor as children's mounts. Hackney ponies are used mainly for show driving.

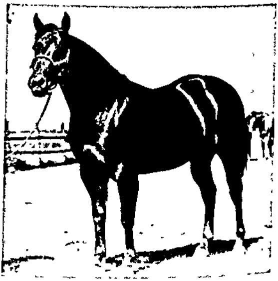


Fig. 4. American quarter horse stallion. (*Photograph by J. A. Stryker*)

Feeding. Individual feeding is the usual practice, each horse being fed according to his use. Good pastures are essential to successful horse production. Mixed grasses and legumes in pastures provide the nutrients required by all horses and are the most natural and economical feed. Horses on pasture exercise freely, remain more content, and show greater mental stability than when closely confined. Idle horses can be maintained satisfactorily on grass during the grazing season and on good hay during the winter months.

However, when horses get strenuous exercise under saddle or in harness, their nutritional needs cannot be met solely by a roughage ration. More concentrated feed must be provided in the form of

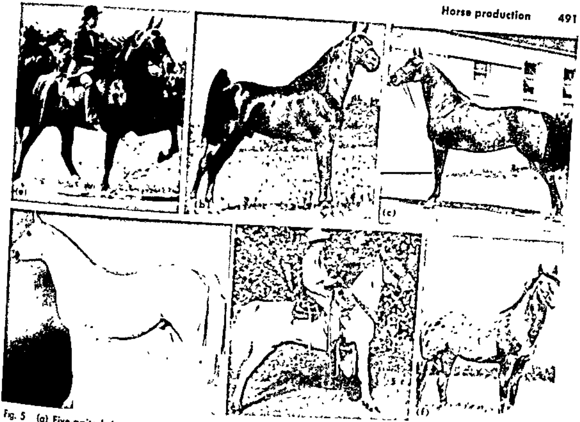


Fig. 5 (a) Five-gaited American saddle horse (photograph by McClasky). (b) Tennessee walking horse (USDA photograph). (c) Morgan mare. (d) Arabian stallion (USDA photograph). (e) Palomino gelding. (f) Appaloosa stallion (photograph by H. H. Sheldon).

grain. The amount of grain fed to horses varies directly with their use. F. B. Morrison, in *Feeds and Feeding* (1956), gives the following recommendations for feeding horses. Idle horses: Chiefly or entirely on pasture and roughage. Horses at light work (1–3 hr/day): .4 to .75 lb grain and 1.25 to 1.5 lb hay per 100 lb of body weight. Horses at medium work (3–5 hr per day): .75–1 lb grain and 1–1.25 lb hay per 100 lb of body weight. Horses at hard work (5–8 hr per day): 1 to 1.4 lb grain and 1 lb hay per 100 lb of body weight.

Oats is the standard grain for feeding horses. Oats and a grass hay, such as timothy, make a balanced ration (*see* TIMOTHY). Corn is often fed, especially in the Corn Belt states, but is a more fattening feed than oats and is lower in protein (*see* CORN). Corn and a legume hay make a balanced ration. Barley, if cracked, makes a good horse feed (*see* BARLEY). Wheat bran, being high in protein and somewhat laxative, is often fed to horses not on pasture. Linseed oil meal may be added to rations low in protein and including hay of poor quality (*see* PROTEIN).

Timothy hay is the standard hay for feeding horses. However, good legume hays, such as alfalfa, red clover, or lespedeza, or mixed grass and legume hays are preferred for feeding brood mares, weanlings, and young growing horses (*see* CLOVER; LESPEDEZA). Such hays are richer in the minerals, vitamins, and proteins needed for growth and reproduction (*see* VITAMIN).

If horses graze on good pasture in the summer and have an adequate balanced ration of grain and high-quality hay during the winter, their mineral and vitamin needs usually are met, except for common salt. This should always be available. In iodine-deficient areas, iodized salt should be fed, especially to pregnant mares, to prevent goiter in young foals. Horses also need 10–12 gal of water per day on the average.

Disease and parasite control. Control measures frequently require the services of a professional veterinarian. However, the practical horseman can prevent some troubles by good management practices. Daily inspection may permit early detection of disease and treatment before the condition becomes serious.

The breathing of a horse should be free, soft, and noiseless. The normal respiration rate for a horse at rest is 8–16 per minute. Normal pulse rate for a horse at rest is 36–40 per minute. The pulse may be taken by palpating an artery that crosses the jawbone just in front of the large cheek muscle. Normal body temperature (rectal) for a horse at rest is 100°F. A horse's coat should be smooth and glossy, and his skin loose and supple.

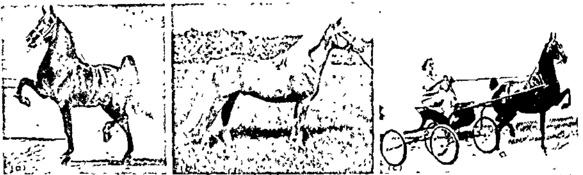


Fig. 6. (a) Shetland pony stallion (photograph by McClasky). (b) Welsh pony stallion (photograph by J. L. A. duPont). (c) Hackney pony mare (photograph by McClasky).

The weight should be borne equally on both front feet. The hindlegs, however, are rested alternately.

Common preliminary symptoms of disease include high temperature; fast, irregular, or noisy respiration; rapid or weak pulse; loss of appetite; sweating without known cause; stiffness; lameness; coughing; inflamed mucous membranes; discharges from nose, eyes, or genitals; diarrhea; constipation; dejection; restlessness; rolling; groaning; heat or swelling in any part.

Internal parasites of horses must be controlled by treatment. Effective prevention of infestation is impossible. Conditions that favor good pasture also favor the completion of the life cycle of horse parasites and the likelihood of infestation.

Strongyles, or blood worms, are the most damaging of the internal parasites of horses (*see* STRONGYLOIDEA). These intestinal worms range from 1/4 to 2 1/2 in. in length. When present in large numbers they cause anemia, weakness, and emaciation. Phenothiazine is commonly used to eliminate strongyles. Many horsemen add 2 g of the drug to the horse's feed each day.

Ascarids, or large roundworms, are particularly harmful to foals and young growing horses (*see* ASCARIDOIDEA). Yellowish-white in color and up to 1 ft in length, ascarids stay in the small intestine. Toxins produced by these worms may kill a foal. Carbon disulfide given by means of a stomach tube eliminates roundworms.

Bots, the larvae of the bot fly, attach to the wall of the horse's stomach and interfere with the normal work of the stomach (*see* ...

... and rotenone are the safest (*see* INSECTICIDE).

Sales. Selling horses profitably is the end and aim of most horse breeders. Young stock must be well bred, well-fed, and well-managed to command high prices. In addition, older horses must give evidence of good training and ability to perform well. Although most horses are sold privately, horse auctions are common throughout the United States. The most noted of these auctions are those for Thoroughbred yearlings held each summer at the Keeneland race track, Lexington, Ky., and at Saratoga, N.Y. Only a very select group of yearlings from the best studs is handled at these sales. During the 1950s the average sale prices per animal have ranged from $6,000 to $10,000. Standardbred yearling sales patronized by the leading market breeders are held each fall at Lexington, Ky., and at Harrisburg, Pa. [J.M.K.]

Bibliography: See AGRICULTURAL SCIENCE (ANIMAL).

Horsepower

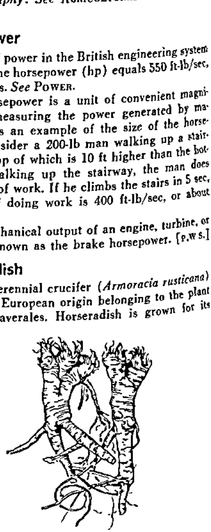
The unit of power in the British engineering system of units. One horsepower (hp) equals 550 ft-lb/sec, or 746 watts. *See* POWER.

The horsepower is a unit of convenient magnitude for measuring the power generated by machinery. As an example of the size of the horsepower, consider a 200-lb man walking up a stairway, the top of which is 10 ft higher than the bottom. In walking up the stairway, the man does 2000 ft-lb of work. If he climbs the stairs in 5 sec, his rate of doing work is 400 ft-lb/sec, or about 3/4 hp.

The mechanical output of an engine, turbine, or motor is known as the brake horsepower. [P.W.S.]

Horseradish

A hardy perennial crucifer (*Armoracia rusticana*) of eastern European origin belonging to the plant order Papaverales. Horseradish is grown for its

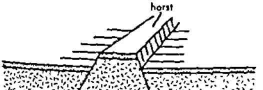


Horseradish, *Armoracia rusticana*.

pungent roots which are generally grated, mixed with vinegar and salt, and used as a condiment or relish. Propagation is by root cuttings and the crop is grown like an annual. The individual roots or sets are uncovered by hand usually twice during the summer and stripped of all side roots. Maliner Kren is a common variety. Harvesting of the roots occurs in the fall, usually 3½–4 months after planting. Production in the United States is limited to northern areas; Illinois, Wisconsin, and Missouri are important producing states. *See* PAPAVERALES; VEGETABLE GROWING. [H.J.C.]

Horst

A segment of the earth's crust, generally long as compared to its width, that has been upthrown relative to the adjacent rocks (see illustration). Horsts range in size from those that have lengths and upward displacements of a few inches to those that are tens of miles long with upward displacements of thousands of feet. The faults bounding a horst on either side commonly have inclinations of



Simple horst. (*From A. K. Lobeck, Geomorphology, McGraw-Hill, 1939*)

50–70° toward the downthrown blocks, and the direction of movement on these displacements indicates that they are gravity faults. These relationships suggest that horsts develop in regions where the crust has undergone extension. They may form in the crests of anticlines or domes, or may be related to broad regional warpings. *See* FAULT AND FAULT STRUCTURES; GRABEN; WARPING, EARTH CRUST. [P.H.O.]

Horticultural crops

Food-producing plants such as tree fruits, small fruits including cane fruits, bush fruits and low-growing berry plants, melons, edible nuts, grapes or vine crops, and vegetables cultivated for their leaves, flowers, fruits, stems, or roots. These plants vary widely in their tolerance to soils, moisture, and temperature, which affects their distribution and adaptation to various localities.

Insects and diseases also greatly influence the profitable culture of many horticultural plants (*see* INSECTA; PLANT DISEASE; PLANT VIRUS). Protection may be given the crops by chemical spraying, by selection of tolerant plants or varieties, by introduction of resistant strains, or by breeding new varieties of resistant plants. *See* BREEDING (PLANT). Desirable traits vary with plants. Important qualities desired include earlier or more uniform maturation, improved skin or shell texture, earlier or brighter color, better flavor, size, and shape, easier harvest, longer edible life, and better storage or shipping qualities.

Many fruit crops are propagated asexually by stem or root cuttings, whereas most vegetable crops are propagated sexually by seed (*see* REPRODUCTION, PLANT). In propagation of some tree fruits and grapes, certain varieties may be grafted onto more hardy or resistant root stocks, thus better adapting them to adverse soil or weather conditions, or pests (*see* GRAFTING OF PLANTS). The trend toward dwarf trees of apple, pear, and peach has encouraged selection and development of dwarfing stocks.

For most horticultural crops, pollination is essential for fruit set. Therefore, varieties in commercial fruit plantings are arranged to ensure maximum cross pollination because some varieties are naturally self-unfruitful, or because the times of blossoming of desirable pollinating varieties do not coincide. Pollination is not known to affect fruit color directly, but in certain species it does affect size, shape, and seed characters of the pollinated fruits. *See* FRUIT GROWING (SMALL); FRUIT (TREE); *GRAPE CULTURE*; *MELON GROWING*; *NUT CROP CULTURE*; VEGETABLE GROWING. [A.F.V.]

Bibliography: *See* AGRICULTURAL SCIENCE (PLANT).

Hot-water heating system

A heating system for a building in which the heat-conveying medium is hot water. Heat transfer in Btu equals pounds of water circulated times drop in temperature of water. For other liquids, the equation should be modified by the specific heats. ... be used for cooling

as radiators, convectors, baseboard radiators or panel coils. A piping system connects the heat source to the various heat-emitting units and includes a method of establishing circulation of the water or other medium and an expansion tank to hold the excess volume of water as it is heated and expands.

Radiators and convectors have such different response characteristics that they should not be used in the same system.

In a one-pipe system, radiation units are by-passed around a one-pipe loop, which should only

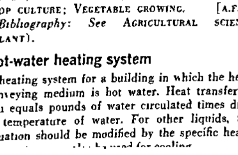


Fig. 1. One-pipe system.

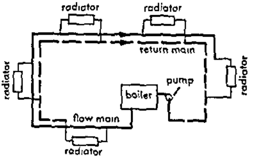


Fig. 2. Two-pipe reverse return system.

be used in very small installations (Fig. 1). In a two-pipe system, radiation units are connected to separate flow and return mains, which may run in parallel or preferably on a reverse return loop, with no limit on the size of the system (Fig. 2).

Circulation may be provided by gravity or pumps. In gravity circulation, each radiating unit establishes a feeble gravity circulation; hence such a system is slow to start, is unpredictable, and is not suitable for convectors, baseboard radiation, or panel coils because circulating head cannot be established, and circulation cannot be supplied to units below the mains. The pipes must be large in size.

In forced circulation, a pump is used as the source of motivation. Circulation is positive and units may be above or below the heat source. Smaller pipes are used

For perfect operation it is imperative that the friction head from the heat source through each unit of radiation and back to the heat source, be the same.

Expansion tanks may be open or closed. Open tanks are vented to the atmosphere and are used where the water temperature does not exceed 220°F (at sea level). They provide the safest operation, practically free from explosion hazards. Closed tanks, used for higher water temperatures, must be provided with safety devices to avoid possible explosions.

One outstanding advantage of hot-water systems is the ability to vary the water temperature according to requirements imposed by outdoor weather conditions with consequent savings in fuel. Radiation units may be above or below water heaters, and piping may run in any direction as long as air is eliminated. The system is practically indestructible. Flue-gas temperatures are low, resulting in fuel savings. The absence of myriads of special steam fittings which are costly to purchase and to maintain is an important advantage.

Hot-water is admirably adapted to extensive central-heating wherein high temperatures and high pressures are used, and to low-temperature panel-heating and -cooling systems. See Comfort control.

[E.L.W.]

Bibliography: American Society of Heating and Ventilating Engineers, *Heating, Ventilating, Air Conditioning Guide*, vol. 37, 1959; F. E. Giesecke, *Hot-water Heating and Radiant Heating and Radiant Cooling*, 1947.

Huebnerite

A mineral with the chemical composition $MnWO_4$. Huebnerite is the manganese member of the wolframite solid-solution series. It commonly contains small amounts of iron. It occurs in monoclinic, short, prismatic crystals. Fracture is uneven. Luster varies from adamantine to resinous. Hardness is 4 on Mohs scale and specific gravity is 7.2. Huebnerite is transparent and yellowish to reddish-brown in color; streak is brown. It is fusible with difficulty. For occurrence, test for tungsten, and use, *see* Wolframite. [E.C.T.C.]

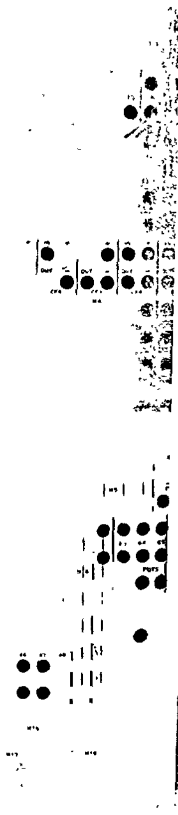
Human engineering

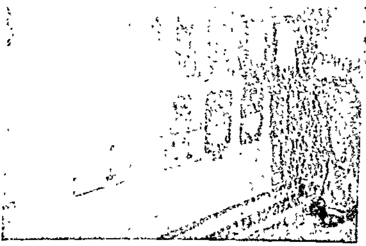
The area of knowledge dealing with the natural capabilities and limitations of the biological man as they relate to machines and systems. Specifically, human engineering can be defined as that field of activity which seeks to ensure that the tools and machines used by man, as well as man's working environment, are compatible with his natural capabilities and preferences. This activity has also been termed human factors engineering, biomechanics, psychotechnology, engineering psychology, applied experimental psychology, and ergonomics. *See* Psychology, physiological and experimental.

As an applied endeavor, human engineering is a professional field primarily linked with engineering. Human-factors data are being used in the design of automobiles, aircraft cockpits, dragline cabs, seats for aircraft, trains, and space capsules, clockface dials, altimeters, navigation maps, submarine controls, communication equipment, control knobs, radar consoles, and other equipment. Work spaces and environments have also been designed to take account of man's sensitivity to such physical factors as light, sound, heat, humidity, cold, vibration, noxious gases, and acceleration forces. The human engineering field will increase as man continues his efforts to explore to greater distances below the earth's crust, to greater depths in the earth's oceans, to greater distances from the earth's surface, and wherever he must take with him a livable environment.

Historical development. The study of animal capabilities is not new. However, studies of man's capability were limited to his physical characteristics, and centuries passed before man's intellectual capacity was recognized as a necessary component of his growing physical power through machines. With the advent of the steam engine and the many mechanical devices developed during and after the Industrial Revolution it became increasingly necessary to rely on human intelligence to plan, guide, and control mechanical efforts.

The significance of the human being as a major factor in man-machine systems was recognized by Francois Dupin in 1829 and later toward the end of the nineteenth century by the engineer Frederick W. Taylor, who demonstrated tremendous increments in productivity at such jobs as pig-iron handling, bricklaying, and industrial inspection by control and direction of workers' efforts. Later

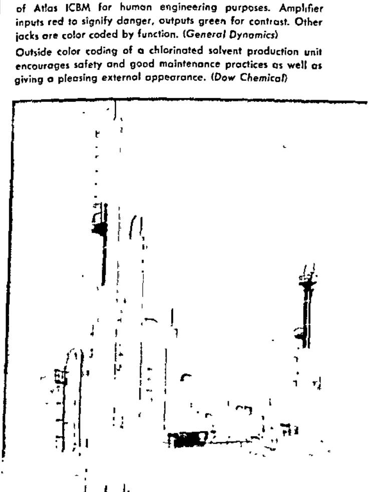




Control board of a chemical processing plant shows actual plant process. Meters are placed at points where measurements are taken Color coding decreases errors and employee fatigue, speeds location and reading of measurements. (*Dow Chemical*)

Color coding on computer switchboard of a dynamic simulator of Atlas ICBM for human engineering purposes. Amplifier inputs red to signify danger, outputs green for contrast. Other jacks are color coded by function. (*General Dynamics*)

Outside color coding of a chlorinated solvent production unit encourages safety and good maintenance practices as well as giving a pleasing external appearance. (*Dow Chemical*)



his time-study approach was expanded by the motion study analysis of the engineer Frank B. Gilbreth, and his wife, Lillian M. Gilbreth, a psychologist. The work of these pioneers may be considered the basis of what is now called industrial engineering, which is akin to the human engineering field. *See* INDUSTRIAL ENGINEERING.

In the early stages of World War II, it became evident that the potential of the advanced war devices created by scientists and engineers often was not realized because of shortcomings of the average military operator. This situation led to a vast training program, the magnitude and urgency of which led to the U.S. Navy's program of "synthetic training," a division of which was described as human engineering. Synthetic training employed training devices to simulate the operation of military equipment under combat conditions. The simulators recreated the real task and attending environment so closely that proficiency or the lack of compatibility between man and machine could be observed readily and with certainty. The shortcomings of devices and their controls became equally evident. These findings encouraged designers of equipment to design their equipment to fit the biological characteristics of the average operator.

Fields covered. Human factors data is generated and applied to equipment and systems design by many scientific and professional specialties. Information is provided by the fields of physical anthropology, anatomy, physiology, psychology, biophysics, and other biological and social sciences. Engineers, architects, and equipment designers are primarily users of the data in their designs, although they may also produce data. *See* ANTHROPOLOGY, PHYSICAL; BIOPHYSICS.

Work-area lighting recommendations*

Task	Illumination, foot-candles
Extremely detailed and prolonged work (inspection of dark yard goods for minute flaws)	100
[illegible]	50
[illegible]	25
[illegible] adaptation is required	5

* From J D Vandenberg and C T Goldsmith, Human factors engineering, *Machine Design*, 1958

The information collected and used by human engineering personnel includes data on sensitivity of the various sense receptors, the characteristics of human movements, dimensions of the body in both static and dynamic states, decision and problem-solving capacities of the machine operator, the capacity of workers to withstand fatigue and monotony, and the effects of drugs and toxicological hazards on work performance In addition, human engineering considers the effects of clothing on work performance, nutritional determinants of worker performance, and the effects of prior training and experience on the worker's efficiency in maintaining and operating novel equipment. It is evident that human engineering must combine the increasingly complex products from the whole spectrum of science with the even more complex human organism. The table illustrates one type of data, light levels used by human engineering personnel in designing work situations.

Efforts in human engineering are directed to such problems as the design of displays and controls,

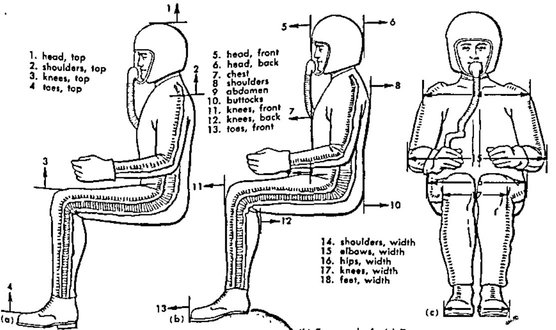


Fig 1. The three dimensions of work-space fit. (b) Fore and aft. (c) Transverse.

the layout of work spaces, and the control of special environmental factors affecting worker performance.

Visual displays. Considerable effort has been expended to present data to equipment operators in a form that is easily and reliably interpreted.

For example, flying a high-speed aircraft "on instruments" (without visual reference) requires reading several instruments simultaneously to obtain a mental picture of the attitude of the aircraft and its speed and direction in both the vertical and horizontal planes. Other instruments must be read periodically to maintain flight and course to destination. Calculation must also be made to effect navigation and to determine speed over the ground, fuel consumption, and remaining range.

The ability to take, comprehend, and react correctly to all these stimuli simultaneously received requires much training and constant practice, because this information supplied to the brain is not in a form that the brain can assimilate naturally. As aircraft speeds increase, the human limits of performance are approached. The estimate of the pilot's mental limits is a human engineering consideration. Even though the instrumentation may be technically adequate, the task of mentally coordinating and visualizing the reading of many instruments may become too great for human reliability.

One solution to the problem is a simple, lightweight electronic computer to collect, assemble and assess the data, and to display the result pictorially. A television picture is produced from the data, portraying a horizon (as in visual flight) and a simulated path along which the pilot flies.

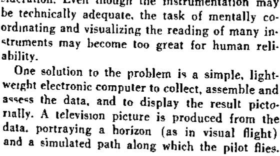


Fig 2. Compatibility of displays and controls. (*From J. D. Vandenberg and C. T. Goldsmith, Human factors engineering, Machine Design, April 17–July 10, 1958*)

The navigational instruments generate a moving mark on a transparent terrain map (or Mercator projection) to give the position of the aircraft at all times. The result of fuel consumption and speed calculations by the computer shows up as a circle of diminishing radius around the aircraft position mark to denote the range remaining.

The pilot thus is provided with an easily comprehensible picture that gives him the necessary current information to fly to his destination. He is relieved of burdensome, time-consuming calculation and can devote his attention to carrying out his mission and coping with any unforeseen circumstances.

Fatigue and emotion. Human engineering must necessarily take into account such indeterminate factors as fatigue and emotions. In general, moderate to high degrees of physical fatigue lead to reduced operator performance. In many work situations operator fatigue can be controlled by careful equipment design, such as optimal force requirements for manipulation of controls, appropriate seat design to support bodily parts, optimal pacing of repetitive hand movements, skillful lighting, and means of relieving monotony. The effects of fatigue can be effaced by strong emotions or excitement caused by such circumstances as combat or competition. Intense emotions resulting from fear or exaltation can radically change man's normal performance. They can make him insensitive or oversensitive to environmental stimuli and pain. They can endow him with exceptional ability or deprive him of his normal skill. Whatever the resulting effect may be, these possibilities must be considered in the design of equipment and the possible effectiveness of the man-machine system. *See* EMOTION.

For instance, student pilots under stress have been known to "freeze" at the controls and exert great strength. The dual controls for the instructor must be designed to overcome the forces so applied, or provision must be made for the instructor to disconnect the student's controls.

Relation to systems engineering. Modern society has created many complex systems in which men and machines interact. Military examples of such systems are combat information centers, air defense networks, gun-crew and missile firing procedures, and the logistic requirements for the movements of large bodies of men, weapons, and vehicles. Civilian examples are found in telephone and other communication networks, post and mail office routines, air traffic control, the stock exchange, ticketing and reservations procedures, inventorying, and production line assemblages. These cases exemplify Paul Fitts' definition of a system as "an assemblage of elements characterized by a common purpose and tied together by an information flow network, the interactions of the elements being of such complexity that it is very difficult to predict the performance of the system from an analysis of the properties of the elements in isolation." This lack of predictability of system performance arises because both the human and machine

elements have their own dynamic characteristics, such as nonlinearities ...

puts to some elements represent the feedback of the outputs from other elements, the final output

systems can be studied only by simulation techniques. See SYSTEMS ENGINEERING.

When faced with a systems problem, the designer must decide whether to use machine elements completely (automation) or whether certain functions can be better allocated to human elements. Such decisions require information (and often further research) on the comparative capabilities of men and machines. Compared to man, machines generally (1) respond to control signals more quickly and accurately (for example, the automatic pilot in routine flight), (2) apply greater force under complete control (such as in hydraulic presses, cranes, and hoists), (3) perform routine repetitive work (they are free from error or decline in output due to boredom and fatigue) (4) can apply and then completely erase stored information (as in a modern computer), and (5) can handle several complex tasks simultaneously (as a multiple packaging machine) Man, on the other hand, can usually surpass machines in such tasks as (1) sensing and detecting small amounts of visual ... d inter- under- words,

(3) improvising during an unforeseen or emergency condition, (4) selectively recalling and making judgments about old information (the wisdom that comes from long experience cannot be easily preset into a machine), and (5) inductive reasoning (that is, man can formulate, invent, discover, or create new principles from empirical data). The purpose of this brief comparison of man and machine is to emphasize the superiority of the human being in dealing with changing situations and unforeseen problems. The basis for man's flexibility is his central nervous system which is composed of some 5×10^{10} nerve cells. It has been estimated that the underlying neurophysiological basis of man's behavior involves the processing of a billion bits of information a second, a rate of data handling far greater than the capacity of any contemporary machine or computer and with incomparably more compact equipment. See INTELLIGENCE; LEARNING THEORIES; MEMORY.

The intelligent design of a man-machine system, therefore, must involve human engineering concepts as well as those directed solely to the arrangement of mechanized components in order to achieve the optimum configuration and the most efficient result. [L.D.F.]

Bibliography: P. S. Allen and E. V. Saul, *Human Factors*, 1:21–49, 1958; A. Chapanis, W. R. Garner, and C. T. Morgan, *Applied Experimental Psychology*, 1949; E. J. McCormick, *Human Engineering*, 1957; Special Devices Center, Office of Naval Research, *Handbook of Human Engineering Data*, 2d rev. ed., 1952; J. D. Vandenberg and C. T. Goldsmith, Human factors engineering, *Machine Design*, April 17–July 10, 1958.

Human genetics

The study of human heredity and variation, including the physical basis (chromosomes and genes), the mechanisms whereby hereditary factors are transmitted from one generation to the next, the processes whereby genes determine characteristics, and the distribution of alternative genes and characteristics in different populations and races and the action of ...

nation of the inheritance pattern, population analysis, twin studies, biochemical genetics, quantitative inheritance, and the inheritance of pathologic traits. For a discussion of human genetics that considers improvement of the genetic endowment of human population by scientifically directed selection see EUGENICS.

The physical basis of human heredity is in all essential respects the same as in other organisms. Each cell contains in its nucleus threadlike chromosomes. Because of the nature of mitotic cell division, these chromosomes, barring some accident, are identical in number and kind with those present in the zygote formed by fusion of egg and sperm, and from which all other cells of the individual are derived. The chromosomes are nucleoprotein, and it is to be presumed that their deoxyribose nucleic acid is, as in other organisms, the actual chemical code of information transmitted through the germ cells from one generation to the next. The germ cells, formed by the meiotic divisions, carry one-half the chromosomes present in somatic cells and primordial germ cells; that is, the mature germ cells carry one chromosome of each kind instead of two. Fertilization restores the diploid chromosome number, each maternal chromosome is consequently matched by a similar paternal chromosome; and the corresponding genes at a particular site or locus in a given pair of chromosomes likewise form a pair. These paired genes, termed alleles, may be identical (homozygous) or different (heterozygous). See MEIOSIS; MITOSIS.

Difficulties and advantages. Classical approaches to human genetics, such as the identification of the mode of inheritance of particular traits, the study of linkage and recombination, and the determination of mutation rates, are fraught with unusual difficulty because of the inability to make appropriate crosses and analyze the outcome. There is also the ever-present doubt as to true parentage which in critical cases must be evaluated by a battery of tests. Nevertheless, human genetics offers certain advantages not provided by the genetics of other organisms. Twin studies can be carried out to show the relative importance of heredity and en-

vironment in various respects and under various conditions, and one twin of a monozygotic pair can be used as a control when the other is subjected to variable conditions. The great attention given to normal human physiological and biochemical characteristics, as well as to pathological conditions, supplies an enormous fund of genetic knowledge when combined with family studies of incidence. As a consequence, far more mutant traits are known in the human species than in any other, even the famous fruit fly, *Drosophila*. The exact nature of inborn errors of metabolism, to use A. E. Garrod's phrase, is not inaccessible to determination, and human biochemical genetics is advancing rapidly. The vast sizes of human populations and the completeness in some places of medical records make it uniquely possible to determine incidences, to calculate the frequencies of mutant genes in the population, and to determine the spontaneous mutation rates of certain genes. Genetic epidemiology is growing rapidly. A widespread anthropological interest in blood groups and other racially distinctive characteristics has provided a fund of genetic information for the study of population genetics and the elucidation of evolutionary changes occurring because of migration and racial intermixture. The use of tissue cultures of human cells to study mutation and biochemical aspects of gene action is just beginning and may well become a major part of human genetics. Eugenics is profiting from the development of heredity counseling and from the analysis of genetic and nongenetic factors affecting fertility and fecundity. The determination of true paternity, or more commonly the exclusion of falsely asserted paternity, is one of a number of important medicolegal applications of human genetics. See BLOOD GROUPS; CULTURE, TISSUE; POPULATION GENETICS.

Chromosome number and linkage. It is now well established that the normal human chromosome number is 46, including 44 autosomes and 2 sex chromosomes. The long-accepted count of 48 was based on older cytological techniques far less accurate than the new tissue culture, smear, and chromosome-spreading techniques which avoid sectioning of cells (Fig. 1). Although normal individuals with 47 or 48 chromosomes may occur, this remains to be confirmed. There is evidence that 47 chromosomes may exist in intersexes or hermaphrodites and mongoloid idiots, having probably arisen through nondisjunction of a sex chromosome or autosome, respectively. Forty-eight chromosomes may also occur in abnormal cases. See CHROMOSOME ABERRATION.

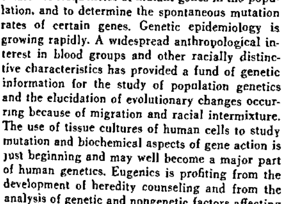


Fig 1. Photograph of human diploid cell with 46 chromosomes, X1950. (E. H. Y. Chu and N. H. Giles, *Human chromosome complements in normal somatic cells in culture, Am. J. Human Genetics, 11(1):63–79, 1959*)

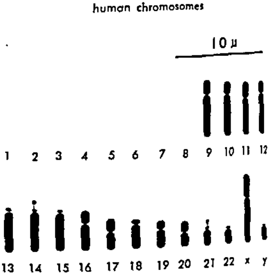


Fig. 2. Drawing of haploid set of human chromosomes. (E. H. Y. Chu and N. H. Giles, *Human chromosome complements in normal somatic cells in culture, Am. J. Human Genetics, 11(1):63–79, 1959*)

A diagram of the haploid chromosome set shows that the 12 largest chromosomes all have median, submedian, or subterminal centromeres (Fig. 2). Among the smaller chromosomes there are 6, including the Y, with acrocentric, that is, nearly terminal, centromeres. The X is a large chromosome with a submedian centromere. Two of the smaller acrocentric chromosomes bear tiny satellites.

There are 23 pairs of human chromosomes; each human gene should therefore eventually be assignable to a particular one of 23 linkage groups; but at present, although much effort has been expended to detect cases of linkage and although elaborate statistical methods have been derived for testing linkage, few autosomal linkages are known. The best established are as follows: Lutheran-Lewis blood groups, 8.3%; nail patella-ABO blood groups, [illegible] blood groups [illegible] such [illegible] have [illegible] very been suspected by many geneticists of [illegible] very closely linked loci, although evidence of internal recombination is lacking. Thus the Rh alleles are widely attributed, following R. A. Fisher and R. R. Race, to variations at three very closely linked loci.

DCE, like pseudoalleles or sites within a functional segment (or cistron) in other species. This view is sharply contested by other immunogeneticists.

Sex is determined in the human species by a particular pair of chromosomes, represented in the female sex by two similar X chromosomes, and in the male sex by one X and one Y (see SEX DETERMINATION). The Y chromosome, possessed exclusively by males, was formerly thought to carry several mutant genes which would be transmitted by a male exclusively to all his sons. The pedigrees on which these conclusions were based have recently been seriously questioned, and the existence of this class of (holandric) genes must be regarded as very dubious. Most of the Y chromosome is homologous with the proximal portion of the considerably larger X chromosome, and crossing over might be expected to occur between the Y and this part of the X, because chiasmata are cytologically visible in this region. Eight or nine mutant loci have been assigned to this homologous portion of the X and Y chromosomes, on the basis of statistical evidence of their partial sex linkage. In order of closeness of linkage with sex, these are total color-blindness, xeroderma pigmentosum, Oguchi's disease, spastic paraplegia, recessive epidermolysis bullosa, retinitis pigmentosa, hemorrhagic diathesis, and convulsive disorders. The validity of this partial sex linkage has also been questioned. See SEX-LINKED INHERITANCE.

The portion of the X chromosome present in two doses in females and in a single dose in males carries genes that show typical sex linkage, and of course must be linked with each other. The best-known case is the linkage between hemophilia and dichromatic (red-green) color-blindness, which in progenies of women heterozygous for both exhibit about 9.8% of recombination (Fig. 3). Besides classical hemophilia, another blood clotting defect, distinguished from it in 1952 and known as Christmas disease, is also sex-linked. The two principal forms of dichromatic color-blindness, protanopia and deuteranopia, are likewise both sex-linked. In neither case is it known whether the similar disorders are allelic or at different loci. Some 25 other sex-linked conditions are known, although it must be emphasized that a number are not clearly characterized and that phenotypically similar or identical conditions may occur by mutation at autosomal loci, that is, loci not on the X or Y chromosome. Myopia and pseudohypertrophic muscular dystrophy are two of the most widespread and important of these other sex-linked conditions. Little is known about the linkage relationships of any of these other loci to the loci of hemophilia and color-blindness. [H.B.G.]

DETERMINATION OF INHERITANCE PATTERN

The first step in all genetic analyses is to determine whether a character is inherited and if so, how it is inherited. Only after this is known can the study of the physiology of gene action and other problems be profitable. The investigator wants to know whether a change in a single gene is sufficient to cause the character or whether more than one gene must be changed. If it is a single gene it is necessary to know whether the gene change is dominant or recessive and whether the gene is on an autosome or on the sex chromosome. Special methods are required in human genetics to determine how a character is inherited. This section is devoted to the analysis of patterns of inheritance of characters resulting from single gene changes.

Individual pedigrees. Patterns of inheritance may be recognized only by the study of families. For genetic purposes, the term family means, as a

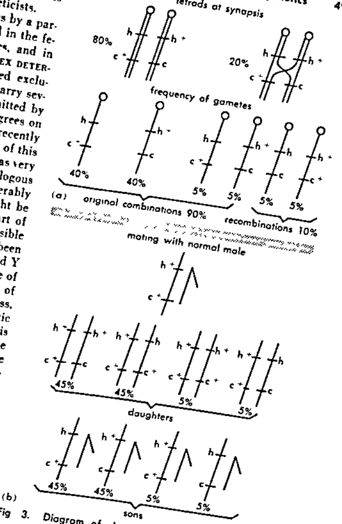


Fig. 3. Diagram of chromosome recombination for hemophilia, h, and red-green color-blindness, c. (a) Crossing over between the sex-linked gene loci of hemophilia and red-green color-blindness in a woman heterozygous at both loci. (b) Among her sons, there is a 10% probability of manifested recombination between the sex-linked loci. An equal proportion of daughters are carriers of crossover chromosomes.

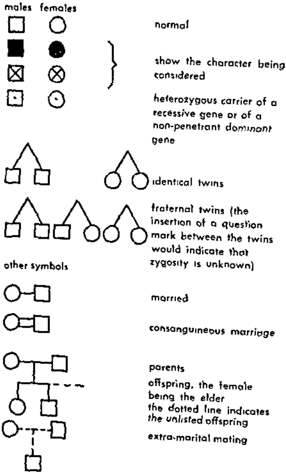


Fig 4. Some symbols in common use *in the United States* for presenting human pedigrees

minimum, parents and offspring; it may, however, include additional relatives.

Various symbols are used to simplify the presentation of a pedigree and hence, also, the analysis of the pedigree. Although no single set of symbols is used universally, those in Fig. 4 are in general use in the United States. Other symbols will be explained as they occur in pedigrees demonstrating simple patterns of inheritance.

Figure 5 shows a three-generation pedigree demonstrating dominant inheritance of a rare characteristic. The inheritance is best described as autosomal dominant because although rare, the character occurred in three successive generations, each affected individual had an affected parent, and an affected male had an affected son, thus excluding sex-linked dominance. Although the frequency of affected offspring among the offspring of affected parents does not depart significantly from 50%, in small pedigrees it is of little or no value to consider the fit of the data to expected frequencies. In evaluating the ratio of normal to affected only the offspring of affected individuals are used; the proband is not counted.

The term proband or propositus designates each individual who has been found simply because he has the character being studied. Those found because the presence of an affected relative led to their examination are not termed probands, but are called secondary cases. These distinctions are important for the methods of analysis of family data, discussed in a later section.

Figure 6 presents a pedigree in which the heterozygote is different from either homozygote. This is usually true for blood group antigens, where the situation is referred to as codominance because

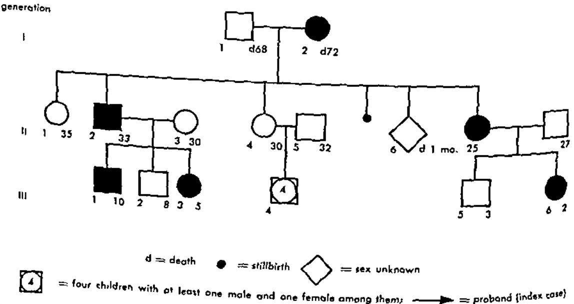


Fig. 5 Hypothetical pedigree of a rare autosomal dominant character showing *complete* penetrance. Roman numerals indicate generations, arabic numerals to the lower left of the *symbols* indicate individuals within a generation; numbers to the lower right indicate ages at the time the pedigree was taken or age at death.

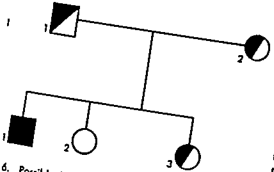


Fig. 6. Possible family pattern when the heterozygote is different from either homozygote. The half-shaded symbols represent the heterozygotes.

each antigen present in the heterozygote reacts at least qualitatively as it does in the respective homozygote. A similar pattern occurs in the case of some hereditary diseases in man, for example thalassemia and sickle-cell disease. In each of these the heterozygote is essentially normal, but can be distinguished from either homozygote by appropriate laboratory procedures. These are cases of incomplete dominance.

Another simple type of autosomal inheritance concerns recessive characters such as albinism. An individual pedigree is not likely to contain sufficient information to permit the conclusion that inheritance of such a trait is recessive. Some features of pedigrees for rare recessive characters are that (1) the parents are usually normal (they are always normal if the character leads to early death, as in cystic fibrosis), (2) collateral and ancestral relatives are rarely affected, and (3) there may be an increased frequency of consanguineous matings. The last criterion is becoming of decreasing value in the United States. First-cousin marriages occur in the general population with a frequency of about 1 in 2000; hence, the expected absolute increase in frequency of consanguineous marriage among the parents of those affected with a rare recessive disease is not great enough to be easily detected. The expected increase in consanguinity is as low as 1.56-fold for recessive characters which occur in 1% of the population; it is 63.44-fold for characters which occur in 0.0001% (1 in 1,000,000) of the population. The former increase (from 1 in 2000 to 1.6 in 2000) requires an enormous sample to detect, one much larger than is necessary to establish the pattern of inheritance by other methods. The latter increase (that is, to 63 in 2000) may be detected in a relatively small sample, but such a sample is difficult to obtain for a character which occurs only 1 time in 1,000,000. Similar arguments hold for characters with intermediate frequencies. It is apparent that first-cousin marriages will only rarely be observed among the parents of those bearing rare recessively inherited characters. Nevertheless, first-cousin marriages, when they are observed among the parents of affected children, lend considerable support to an hypothesis of recessive inheritance. It remains true that recessive inheritance may best be recognized by analyzing data from numbers of two-generation pedigrees. This will be discussed in a later section.

Pedigrees demonstrating sex linkage are now examined. Figure 7 presents a condensed version of a large pedigree of a rare disease which is fatal in childhood and which is inherited as a sex-linked recessive. The pedigree also illustrates the common practice (when dealing with rare characters not due to an autosomal recessive mutation) of omitting from the pedigree all those not genetically related to the individual known or believed to have introduced the character into the pedigree. It demonstrates also another way of indicating affected individuals.

Distinguishing features of pedigrees showing recessive sex linkage are that (1) affected males do not transmit the character (the gene) to their sons, (2) normal males have only normal descend-

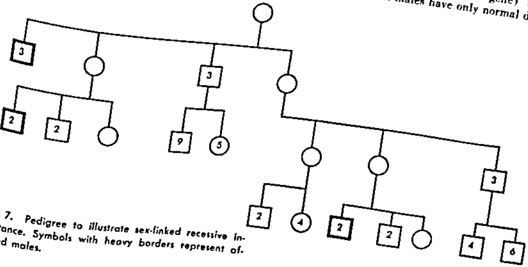


Fig. 7. Pedigree to illustrate sex-linked recessive inheritance. Symbols with heavy borders represent affected males.

ants, (3) affected males other than sibs are related to each other via females, and (4) affected females do not occur. Exceptions may occur if a male in such a pedigree marries a carrier (heterozygous) female. In such a mating, affected males may have affected sons (the gene coming from the mother) and affected daughters, and normal males may have affected daughters. The same features hold for sex-linked dominant characters. In addition, however, all daughters of affected males will be affected and normal females will have only normal descendants.

Methods are available for the quantitative evaluation of the agreement of the observed number of affected with the expected number.

Medicolegal uses of genetics. Only genetic factors which are widespread in the population and which are inherited by the way of a simple, well-understood, and adequately tested genetic mechanism are useful for medicolegal purposes. At present the only characters which fulfill these requirements are the ABO, MN, and Rh blood group systems. Accordingly, they are the only ones universally accepted for medicolegal use.

Genetic data concerning the blood groups and types are used in cases of disputed paternity, parentage, or identity In each of these situations the genetic data may be used to exclude the claimed relationship, but not to establish it, because of the existence of many individuals of the same blood groups and types, any one of whom, from the genetic viewpoint, could be substituted for each of the others.

In cases of disputed paternity the question to be answered is whether this child of this mother could have this man as its father. If the child has an antigen not present in either its mother or alleged father, the man is excluded as the father; similarly, if the child does not have an antigen which the alleged father for genetic reasons had to transmit to his offspring, the man is excluded as the father. Tables 1 and 2 present in summary form the ABO and MN blood groups of men who could not have fathered the child in a given mother-child pair. A similar table may be devised for the Rh system.

By use of the three blood group systems, slightly over 50% of men who are wrongly accused of fathering a child may be shown not to be the father.

The situation is genetically more complicated for the detection of falsity in claims of parentage. Here the basic question is whether the given individual could be the child of the given couple. The following example illustrates this point. An individual of group M claims as his parents a woman who is group M and a man who is group N; since this couple can have only MN [illegible]

groups is to test for possible exchange of infants. Finally, there is the use of blood groups for identification purposes. Exclusion results when the sample fails to match the individual's blood exactly.

Table 1. ABO blood group of men excluded as fathers of a given mother-child pair

Mother's blood group	Child's blood group			
	O	A	B	AB
O	AB	O, B	O, A	*
A	AB	None	O, A	O, A
B	AB	O, B	None	O, B
AB	*	None	None	O

* Mother-child combination does not occur

Table 2. MN blood type of men excluded as fathers of a given mother-child pair

Mother's blood type	Child's blood type		
	M	MN	N
M	N	M	*
MN	N	None	M
N	*	N	M

* Mother-child combination does not occur.

By 1959, 13 of the 50 states had passed laws authorizing courts to order blood tests in cases of disputed paternity. Efforts are being made to have all the states adopt a uniform law on blood tests in cases of disputed paternity.

Recessive inheritance ascertainment. Data for studies in which recessive inheritance is suspected usually consist of a series of families composed of parents and offspring. Families involving recessive inheritance are ascertained (found) via affected children. The pattern of finding the affected children influences the distribution of the families in the sample; hence, the method of analysis to be used must take into account the method of ascertainment.

Three patterns of ascertainment may be recognized for recessive inheritance.

Truncate selection. Every individual in a community is examined regardless of his condition or that of his relatives. Hence it is certain that any affected individual will be independently ascertained ($\pi = 1$, where π is the probability of being independently ascertained, that is, of being a proband). As a result, all families with at least one affected child are equally likely to be found. The distribution of families of a given size s with one or more affected children is equivalent to a binomial distribution with one term (no children affected) missing, that is, a truncated binomial distribution.

Multiple incomplete selection. The probability of independently ascertaining an affected person, that is, of finding him only because he is affected, is less than unity, but is great enough that a family with two or more affected children may be ascertained more than once. The probability of ascertaining a family is thus a function of the number r of affected persons in the family and of the probability of ascertaining an affected individual π.

Single selection. Each family is ascertained only once regardless of the number of affected individuals in the family. Hence, the probability of as-

certaining an affected individual is so low that it approaches zero ($\pi \rightarrow 0$).

Statistically powerful methods which take into account the effects of mutation, phenocopy, and incomplete penetrance are being developed. These methods will be of greatest value to the specialist.

The methods suitable for truncate selection and for single selection are older, simpler, and less powerful, but useful. For truncate ascertainment, an estimate p_e of the proportion p of recessives in samples is

$$\frac{R/pq - \sum_s Bn_s}{\sum_s Wn_s}$$

and the variance is $1/\Sigma Wn_s$; for single selection, p_e is $(R - N)/(T - N)$ and the variance is

$$\frac{p_e(1 - p_e)}{T - N}$$

where p_e = estimate of p
p = assumed value of p
$q = 1 - p$ = assumed value of q
s = size of family
n_s = number of families of size s
R = total number of affected in sample
$B = (s^2pq^{s-2})/(1 - q^s)^2$
$W = s(1 - q^s - spq^{s-1})/pq(1 - q^s)^2$
N = number of families in sample
T = total number of children in sample

There is no simple method for multiple selection. Such data had better be given to a specialist for analysis.

It is necessary to group the data by families according to the number of affected parents per family and to analyze each group separately.

Maximum likelihood. The method of maximum likelihood is the method of choice to estimate the frequency p of affected offspring among the offspring of groups of families. If the data have been obtained by truncate selection p may be estimated from the following equation, the terms of which are explained above.

$$R = p_e \sum \frac{sn_s}{1 - q_e^s} \qquad (1)$$

The variance of this estimate is

$$\frac{p_eq_e(1 - q_e^s)^2}{\Sigma s(1 - q_e^s - sp_eq_e^{s-1})n_s} \qquad (2)$$

The calculation of these values requires a preliminary estimate or guess of the value of p and the evaluation of the right-hand side of Eq. (1) to determine how well it approximated R. If the approximation is poor, another value of p is chosen, and so on until a good fit to R is obtained. The value of p yielding this fit is p_e, the desired estimate of p. Such a method of analysis is called an iterative method. The derived estimate of p is substituted in Eq. (2) to calculate the variance. The mechanics of the calculations required for this method are formidable. D. J. Finney has derived an algebraic equivalent of Eq. (1) which permits the use of tables to estimate p and its variance. The use of these tables (D. J. Finney, the truncated binomial distribution, *Ann. Eugenics*, 14:319–328, 1947–1949) reduces the calculations to a few simple multiplications and additions. Iteration is also required with this procedure.

The maximum likelihood estimate of p for data collected by single selection is straightforward: subtract the observed number of families of the given type from the total number of affected offspring and divide this difference by the difference between the total number of offspring in the sample and the number of families in the sample. As before, p_e is compared with an appropriate theoretical value by means of the standard error.

Procedures suitable for various special situations are available. [A.C.S.]

POPULATION ANALYSIS

A population, from the genetic viewpoint, constitutes a pool of genes with certain frequencies. The system by which the individuals mate determines how these genes are combined to form the genotypes. The analysis deals first with the mating system, the existence of genetic polymorphism, and the genetic meaning of racial differences. Some of the factors that tend to change the content of the gene pool are also discussed.

Mating systems. Two mating systems, panmixis or random mating, and assortative or nonrandom mating, are discussed in the following section.

Panmixis. In this system of mating, also known as random mating, each member of one sex in a population is equally likely to mate with any member of the opposite sex. Such a population is said to be panmictic. A mating is then a random event, determined by chance. The frequency of certain types of matings is consequently determined by the frequency of the types of individuals in the population. The randomness of matings, however, refers only to certain characteristics of the mating individuals. The matings may be at random with respect to one characteristic (for example, blood groups) and yet not so with respect to some other characteristic (such as height). Random mating is not an absolute term and has meaning only with respect to specified characteristics of a given population. The term is applicable to human as well as to animal and plant populations. The genetic consequence of such a system of mating has been extensively studied because many populations in nature, including man, practice random mating at least approximately with respect to many genetic characteristics. The result of panmixis in a large population with respect to one pair of autosomal genes (locus A, a, for example) is known as the Hardy-Weinberg equilibrium.

Hardy-Weinberg law. This law concerns the stability of the three genotypic (AA, Aa, aa) propo[rtions] in a large panmictic population in the

of disturbing factors such as mutation, selection, migration, and sampling fluctuation. If a fraction p of all the genes at the A-a locus in the population is allele A and the rest ($q = 1 - p$) is allele a in both sexes, then a population consisting of $p^2(AA)$, $2pq(Aa)$, and $q^2(aa)$ will retain the same genotypic proportions in subsequent generations under the system of random mating. Such a population is said to be in an equilibrium state (of a neutral nature). The equilibrium law reflects the conservativeness of a Mendelian population.

Assortative mating. In the broad sense, assortative mating includes all systems of mating other than random mating. In the narrow but usual sense, it excludes the various systems of inbreeding (mating according to genetic relationship) and refers to nonrandom mating with respect to phenotypes. Positive assortative mating means that individuals of the same type tend to mate to a greater or lesser degree more often than they would by chance alone, and the matings between individuals of different types occur less often than dictated by chance alone. A preference based on phenotypic resemblance is involved. Familiar examples in man are the positive correlations between spouses in such characteristics as color, stature, physique, and intelligence level, all of which are controlled to a certain extent by heredity. Positive assortative mating leads to greater parent-offspring correlation. If the degree of assortativeness in mating remains the same for a large number of generations, the population will approach an equilibrium condition. Moreover, the genotypic composition of an equilibrium population reached through assortative mating may be quite different from that reached through an inbreeding system. Negative assortative mating means that the individuals of unlike phenotypes tend to mate more often than expected by chance alone. Continued negative assortative mating of the same degree also leads to an equilibrium condition.

Balanced genetic polymorphism. This is the coexistence of two or more alleles of a locus at substantial frequencies in an equilibrium population. An equilibrium population will consist of various genotypes (with respect to the specified locus) in substantial proportions in spite of their possible differential genotypic selective values. Genetic polymorphism is to be distinguished from polymorphism in which the existence of a *rare allele* in a population is supported by recurrent mutations. The most familiar examples of genetic polymorphism in man are the various blood *groups*, although the selection scheme that maintains such a polymorphism is not yet fully known. For two alleles with the three corresponding genotypes, a superior genotypic selective value (whatever the reason for it) of the heterozygote over the two homozygotes will result in a balanced genetic *polymorphism*. If the selective values of the genotypes AA, Aa, aa are $1 - s$, 1, $1 - t$, respectively, then the equilibrium panmictic population will have the gene frequencies $p(A) = t/(s + t)$ and $q(a) = s/(s + t)$. In man, for example, the individual homozygous for the sickle-cell gene suffers from a severe type of chronic anemia and usually dies young; he has only a small chance to reproduce. The heterozygous individual exhibiting the sickle-cell trait without developing anemia is more resistant to falciparum malaria than the homozygous normal individual. Thus, the heterozygote has a superior selective value over both homozygotes, so as to result in a genetic polymorphism with respect to the sickle-cell locus. The proportion of individuals with the sickle-cell trait may be as high as 20–40% in certain African populations.

Racial differences. All men belong to one and the same species, *Homo sapiens*. The major divisions of man such as Caucasian, Mongolian, Negro, and so on are called races. However, within each major division, there are many minor divisions which are also referred to as races. The difference between two racial groups is usually found not in one but in many biological characteristics. A race, from the genetic viewpoint, is an aggregate of individuals who form more or less a mating group among themselves so that the genic content of this group differs from that of other mating groups. The fact that a marriage even between members of two different major races produces normal children who in turn also produce normal children shows that the chromosomes of all men are homologous. Hence, the racial differences must be attributed to the alleles of the various loci of the corresponding chromosomes. A particular allele may be present or absent in any race, but usually it occurs in all races. The racial differences arise from the differences in allelic frequencies.

The difference in allelic frequencies is a quantitative, not a qualitative one. For example, there are 61% blood group O individuals among the Ethiopians, 46% among the English, and 31% among the Chinese. However, all of these group O individuals have a genotype consisting of the same genes at the blood group locus. For each characteristic the racial differences are due to the differences in the respective allelic frequencies; racial difference as a whole is due to the differences between the various racial groups in the gene frequencies at many loci.

Estimation of mutation rate. Each sort of gene mutates at a certain rate per generation. Some genes are more mutable than others. The natural rate of mutation per locus per generation is very low, being generally of the order of 1 in 10^4 to 1 in 10^5. The mutation rates for human genes, even for the well-known genes, cannot be estimated from directly observed data except in very special cases. The few estimates that have been made are all based on indirect evidence and are subject to large error. At best, they can give only a rough idea as to its order of magnitude. For dominant lethal genes, the mutation rate is equal to one-half the incidence (frequency of new cases) among the new births or among those of the same age group, depending on the time of onset of the lethal trait. For

recessive genes, the mutation rate may be indirectly estimated as $\mu = sq^2$ where μ is the mutation rate to the recessive gene, s is the selection coefficient against the recessive trait, and q is the equilibrium frequency of the recessive gene in the population. The indirect method of estimating the mutation rate is always based on the selection scheme in a population at equilibrium.

Gene flow. Gene flow results from the intermixture of two or more racial groups. Intermarriage implies the exchange of genes between racial groups and thus tends to render their genic content more alike in frequency. The most conspicuous example of racial intermixture in modern history is that between the white and the Negro in the United States. Because the hybrid individual is regarded as a Negro rather than a white according to social convention, this results in an one-way flow of genes from the white to the Negro population. The accumulation of "white" genes in the Negro population increases as intermixture continues. From analyses of the various gene frequencies it has been estimated that the American Negro population contains 25–35% of genes from the whites. The flow of genes in the reverse direction (that is, from Negro to white) is statistically negligible.

Genetic drift. When a population is very large, its gene frequencies remain practically constant from generation to generation, that is, within the limitation that the environmental conditions also remain constant. However, when the population is small, the gene frequencies fluctuate (change in a random manner) from generation to generation even though the environmental conditions remain the same. This random fluctuation (not a systematic change) of gene frequencies is known as genetic drift, and was first studied by S. Wright. The offspring generation is determined only by the actual parents (those who have actually contributed genes to the next generation). The random drift arises from the fact that each generation is the result of the union of a random sample of gametes produced by the parent generation. The smaller the sample of gametes, the larger is the sampling fluctuation in the gene frequencies. As long as a gene frequency is between 0 and 1, the random drift process continues. As soon as it reaches 0 or 1, the gene becomes extinct or fixed in the population and, barring new mutations, there will be no further changes. The genetic drift in small populations [illegible] gosity with respect to all loci in small populations is therefore the result of a stochastic process and is not necessarily correlated with any selective significance.

The effective population size is that portion of the population that contributes to the genetic composition of the next generation. This number is usually smaller than the total number of individuals in the population because not every individual reaches sexual maturity and mates. Nor do [illegible] mate necessarily leave offspring that survive and mature in the next generation. In an ideal population of N breeding individuals, one-half of whom are males and one-half females, the variance of gene frequency q in each generation is $q(1-q)/2N$ and the rates of fixation and of extinction are both $1/4N$ per generation, the total rate of $1/2N$ being known as the rate of decay of variability. Using this idealized population as the standard, the variance of gene frequency or the rate of decay of various other populations may be compared. The actual breeding size (number of male and female parents) of the other populations is to be reduced to an equivalent number N_e so that the variance of q is equal to $q(1-q)/2N_e$ or its rate of decay is equal to $1/2N_e$. Then this calculated number N_e is called the effective size of the population. Factors that tend to diminish the effective size are inbreeding, differential fertility of mating couples, unequal number of male and female parents, and periodic or cyclic depletion in the breeding size. In the case of depletion in breeding size, the effective size is much nearer to the smallest breeding size in the cycle than to the largest size. The effective size of some human populations may be as small as a few hundred or less. [C.C.L.]

QUANTITATIVE INHERITANCE

Human quantitative inheritance has been studied statistically since the latter part of the nineteenth century. However, despite the wide acceptance of Mendelian inheritance for discrete characters, there remained a wide belief that quantitative traits were inherited by some other mechanism. The reso[illegible] the observed correlations in physical measurements between relatives correspond very closely with expectations based on the Mendelian theory of inheritance. It was necessary only to postulate that a large number of genes were involved.

Skin-color inheritance in marriages involving Negro and white persons can be described to a fair first approximation by assuming two pairs of genes, each without dominance and contributing equally in amount of pigment. However, it is certain that the true situation is considerably more complex and that several genes are involved, along with environmental influences.

Physical measurements such as height and weight appear to be inherited in ways at least as complicated as those determining skin color. The number of genes is large enough, and their interaction with each other and the environment complex enough, that there is little hope for studying the effects of individual genes. It is therefore necessary to rely on statistical procedures assuming a large number of genes.

There is great difficulty in separating the effects of heredity from those of the environment. Ordinary family studies do not do this, because members of a family have similar environments as well as similar genes and the two effects are confounded.

The most widely used method of trying to distinguish genetic from environmental effects is by using identical twins who, having developed from a single fertilized egg, have identical heredity. Whatever differences such twins possess must be attributed to environmental differences. However, this does not completely solve the problem, because identical twins may have more similar environments than other twins.

Intelligence, as measured by various tests, has been extensively studied by the twin method. Although it is not possible to draw rigorous conclusions, these studies suggest that when errors of measurement are taken into account, genetic factors may be just as influential in determining intelligence test scores as in determining physical measurements. Studies on personality traits have been still less conclusive, perhaps because of the lesser reproducibility of the testing methods.

[J.F.C.]

TWIN STUDIES

Twin studies comprise a quasi-experimental research technique used particularly in human genetics to explore the interaction of heredity and environment in determination of normal or pathological characteristics. Although the act of bringing forth two animals or children at a birth had always aroused the interest of scholars and artists, it was F. Galton who in 1875 introduced twin studies as a tool for genetic analysis. Based on the regular occurrence of two genotypically different types of twins, one-egg (monozygotic or MZ) and two-egg (dizygotic or DZ), the method called for adequate criteria for zygosity classification (Fig. 8). The standardization of this procedure was long in coming.

Population twin rate. This approximates 2%, with the proportion of same-sex pairs representing about two-thirds of all twins. The 2:1 ratio of same-sex to opposite-sex pairs remains virtually unchanged in all age groups, whereas the twin rate (2.19% of all babies born in the United States since 1928) is reduced to about 1.9% by excess twin mortality within the first year of life. Another check on unbiased sampling in twin studies is provided by the proportion of MZ pairs, using Weinberg's differential method. Weinberg's differential method is used for estimating the proportion of one-egg pairs in a sample of twins that includes male, female, and opposite-sex pairs. A certain proportion of the same-sex pairs is expected to be monozygotic. This proportion is obtained by multiplying the number of opposite-sex pairs by 2 and by subtracting the given figure from the total number of twins in the sample.

Modern similarity method. This method uses dermatoglyphic and blood-type comparisons to determine the zygosity of same-sex twins. DZ twins may be of the same or opposite sex, whereas MZ twins are always of the same sex. In cases when all other diagnostic criteria are indecisive, reciprocal skin grafts may be performed (Fig. 9).

Fig. 8. Twins. (a) One-egg twins; (b) two-egg twins.

The fetal-membrane method is no longer resorted to, because not all MZ pairs are born with one placenta. All monochorial placentae belong to MZ pairs, but at least 20% of MZ twins are dichorial (have two chorions). Hence, about 10% of all dichorial placentae belong to MZ rather than DZ pairs. This is illustrated in the following example. Because two-thirds of all twins are DZ, and one-third MZ, in a sample of 100 pairs of twins, 33 MZ and 67 DZ pairs may be expected. Of the 33 MZ pairs, 7, or 20%, would be dichorionic. Thus, a total

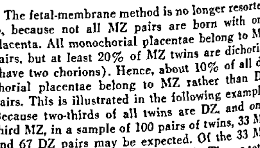


Fig. 9. Skin grafting in twins. (a) One of a dizygotic pair; (b) one of a monozygotic pair. Number indicates postoperative day. (Photograph by Blair O. Rogers)

of 74 pairs will be found to be dichorionic if the 7 MZ pairs are added to the 67 DZ pairs which are all dichorionic. The 7 dichorionic MZ pairs constitute slightly less than 10% of all dichorial placentae.

Twins cannot be monozygotic (MZ) if the blood groups are different, but they may or may not be of the MZ variety if the blood groups are the same. Accuracy of the hematological analysis requires that, in same-sex pairs similar in respect to the major ABO and Rh types, blood typing be continued until a difference appears or until all available antisera have been tried.

Although fingerprints conform to one of three genetically determined types (whorls, loops, and arches), the dermatoglyphic analysis covers both qualitative and quantitative aspects. Quantifying techniques include Wendt's individual pattern score, the sum of homolateral ridge-count differences, and the scores obtained with Slater's discriminant function. Once zygosity is established, twin studies can be organized in various ways, depending on the frequency and type of the trait to be investigated (Fig. 10). See FINGERPRINT.

The original version of the method limits the comparison of phenotypic similarities and dissimilarities to twins in their usual environments. It calls for a representative series of MZ and DZ pairs, whether reared together or apart. No type or degree of variation can be adequately studied with this method unless twins present tangible evidence of a well-defined trait, especially in pathological conditions when the principles of the proband method have to be applied.

Because generalized conclusions cannot be drawn from observations in a few pairs, appropriate sampling procedures with complete ascertainment of twin index cases (rather than pairs) in a certain district are essential. Representative twin data on normal personality characteristics are expressed in terms of varying degrees of intrapair similarity, whereas those on the frequency of pathological traits in cotwins are computed in terms of concordance-discordance. In a sample of twin probands, a pair is called concordant when the cotwin of the index case (proband) shows evidence of the same condition as that ascertained in the index

Fig. 10. Corresponding fingerprints of MZ twins.

case. Discordance refers to the absence of the given pathological trait in the cotwin.

The cotwin-control method is a procedure whereby observational or experimental data are obtained from a few selected MZ pairs whose aptitudes, metabolic reactions, or potentials for adjustment can be compared under different conditions, or in response to planned differences in management. In deliberately dissimilar or spontaneously discordant pairs (one married and one unmarried, or one deaf and one hearing partner), longitudinal comparisons of adaptational patterns, biochemical tests, or psychometric scores may prove illuminating as to etiology, diagnostic classification, therapeutic effectiveness, or personality assessment, always provided the given twins are monozygotic. In a longitudinal study, repeated comparisons (between two twins or two siblings) are made on the same research subjects at different ages rather than on different groups of subjects at certain age levels (cross-sectional studies).

The twin-family method extends its comparative scheme to complete sibships of twin index cases and their parents. The six sibship groups to be compared are one-egg twins, two-egg twins of the same sex, two-egg twins of opposite sex, full sibs, half sibs, and stepsibs. The main advantages of this procedure are its effectiveness as a sampling method and its inconspicuousness as a pluridisciplinary approach to total population surveys which require combined cross-sectional and longitudinal investigations in a cooperative family setting. Whenever necessary, K. J. Holzinger's h^2 values (formula: variance of DZ twins minus variance of MZ twins, divided by the variance of DZ twins) are also obtainable from such family data. Holzinger's h^2 values have been used to demonstrate the role of heredity in intelligence. Intelligence quotients of one-egg and two-egg twins are compared. The resulting h^2 is a quantitative expression of the intrapair variance attributable to genetic factors. [F.J.K.]

BIOCHEMICAL GENETICS

The study of inherited variation in human biochemistry began in the early years of the present century with the work of A. E. Garrod, who was interested in a group of apparently inherited rare diseases in each of which a characteristic disturbance occurred in the normal pattern of the metabolic processes. He called such disturbances inborn errors of metabolism and suggested that in each condition the typical biochemical and clinical findings all could eventually be traced back to a block at some particular point in intermediary metabolism. He thought that this metabolic block could be attributed to a congenital absence of the specific enzyme normally necessary for the catalysis of the biochemical step in question.

Examples. A classical example is alkaptonuria, a condition characterized by the continuous excretion in the urine of large amounts of homogen acid. Homogentisic acid is a normal intermedi

the oxidation of the aromatic amino acids phenylalanine and tyrosine. In alkaptonuria, the enzyme homogentisic acid oxidase necessary for the conversion of homogentisic acid to maleylacetoacetic acid is not present, and so the homogentisic acid which is being continuously formed cannot be broken down any further. Therefore, it accumulates and is excreted in large amounts in the urine. Another example is albinism, a disorder in which there is a deficiency of the enzyme tyrosinase. This enzyme is necessary for the conversion of tyrosine to the pigment melanin. Affected individuals are therefore almost completely devoid of pigment in the skin, hair, and eyes, where melanin is normally deposited.

Implication. An important implication of Garrod's concept of the inborn errors of metabolism was that genetical factors could be important in determining enzyme formation. This idea was later to be greatly elaborated and extended, and in the form of the "one gene–one enzyme" hypothesis emerged as a central working hypothesis in genetics. As applied to the inborn errors of metabolism this hypothesis suggests that the normal allele of the abnormal gene which causes the disease in question is necessary for the synthesis of a particular enzyme. The abnormal or mutant gene cannot perform this function; individuals who carry it in double dose, that is, are homozygous, fail to make the enzyme, and consequently exhibit a specific block in metabolism. This may in turn give rise to diverse clinical consequences. Individuals who are heterozygous, that is, who carry both the normal allele and its abnormal or mutant form, will be capable of making the enzyme. However, they may do so only in limited amounts, and whether or not this leads to clinical disturbances will depend on the quantities of the enzymes actually formed, and on the immediate metabolic requirements. Individuals who suffer from the disease phenylketonuria, for example, are homozygous for a particular abnormal gene. They have a severe disturbance in phenylalanine metabolism as a result of a deficiency of the enzyme phenylalanine hydroxylase necessary for the conversion of phenylalanine to tyrosine. One of the main consequences of this is a severe degree of mental retardation, so that most of the affected individuals are imbeciles or idiots. Their parents, who must be heterozygous for the gene, are not mentally retarded and appear quite healthy. However, they have been shown to have a minor disturbance in phenylalanine metabolism which is qualitatively similar to, but much less in degree than, that found in their affected offspring. This is evidently the result of a partial deficiency of phenylalanine hydroxylase. The enzyme [illegible]

[illegible] deficiency of a particular enzyme but to the formation of an unusual enzyme with properties qualitatively different from those of its normal counterpart. An example of this has recently been discovered in investigation of the enzyme pseudocholinesterase, which is present in the blood plasma of individuals who are excessively sensitive to the drug succinylcholine. Succinylcholine is a muscle relaxant often used in anesthesia and it has been observed that about 1 in 3000 persons suffers a much more prolonged respiratory paralysis than do other people when the drug is administered to them. Succinylcholine is broken down rapidly by normal pseudocholinesterase, and it is thought that this limits the period of activity of the drug in normal people. Succinylcholine-sensitive individuals have in their plasma an atypical form of pseudocholinesterase which differs from the normal enzyme both in its substrate specificities and its inhibition characteristics. Such individuals are homozygous for a particular abnormal gene. The heterozygotes who carry both the unusual gene and also its normal allele have in their plasma both the atypical enzyme and also its normal counterpart in about equal amounts. They occur with a frequency of about 3–4% in the general population.

Synthesis of nonenzymatic proteins. All enzymes are proteins, but not all proteins function as enzymes. If inherited differences in enzyme formation occur, inherited differences in the synthesis of nonenzymic proteins can also be expected. In the last few years many examples of this have in fact been discovered, and their analysis is proving to be of fundamental significance.

The most extensively studied of these are the genetically determined variants of hemoglobin, the red blood cell protein. Their discovery arose from a biochemical investigation of the so-called sickle-cell phenomenon. The red blood cells of certain individuals have the peculiar property of undergoing an alteration in shape when the partial presence of oxygen to which they are exposed is reduced. They change from the normal circular biconcave form to elongated, crescentic, or sickle-shaped forms. Most people whose erythrocytes show this peculiarity are quite healthy and are said to have the sickle-cell trait. However, a small proportion are severely anemic and are said to have sickle-cell anemia. The peculiarity is inherited, and it has been shown that individuals with the sickle-cell trait and those with sickle-cell anemia are respectively heterozygous and homozygous for a particular abnormal gene.

It was discovered that the hemoglobin present in the red cells of patients with sickle-cell anemia differed in its physicochemical properties from normal hemoglobin. Among other things, its solubility when deoxygenated is very much less than normal deoxygenated hemoglobin. This property is responsible for the sickling phenomenon, because when the oxygen tension is reduced the abnormal hemoglobin tends to come out of free solution and so distorts the shape of the red blood cell.

Individuals with the sickle-cell trait were found to have in their red cells a mixture of normal he-

moglobin and the abnormal or sickle-cell hemoglobin. Thus it appeared that the normal gene at the particular chromosomal locus was concerned with the synthesis of normal hemoglobin, and its mutant allele led to the synthesis of the abnormal type of hemoglobin. In the heterozygote where both genes are present, both types of hemoglobin were formed (see Fig. 4).

Hemoglobin is a protein with a molecular weight of about 67,000. It is made up of two identical half molecules, each of which contains about 300 amino acid residues. These are arranged in two polypeptide chains. Electrophoretic comparisons of the abnormal hemoglobin present in sickle-cell anemia with normal hemoglobin reveal that the two molecules differ structurally in only one respect. A particular glutamic acid residue at one point on one of the polypeptide chains in normal hemoglobin is replaced by a valine residue in sickle-cell hemoglobin. In all other respects the molecules appear to be identical.

A single gene difference such as that between the sickle-cell gene and its normal allele is presumably the result of a single mutational step. This is the smallest unit of inherited variation and it may eventually lead to the smallest possible unit of structural difference in the hemoglobin formed, namely, the substitution of a single amino acid residue by another. A change as subtle as this in the hemoglobin molecule can apparently result in an alteration in its physicochemical properties, sufficient to cause major pathological and clinical disturbances. An analogous change in the fine structure of an enzyme protein might well produce an alteration in, or a complete loss of, its enzymatic properties. This could be the underlying basis of the kinds of metabolic upsets characteristic of the inborn errors of metabolism, and quite possibly forms the foundation for human biochemical variation in general. [H.H.]

INHERITANCE OF PATHOLOGIC TRAITS

This section discusses the patterns of inheritance of pathological traits, the etiological interplay of genes and environment, variation in genetic disease, various syndromes, sporadic cases, methods of testing genetic hypotheses, and disorders which result from chromosomal aberrations.

Definitions. The term congenital means merely that manifestations of a trait are present at birth. The trait may or may not be primarily hereditary in etiology. Even though the genetic propensity to a disease is present at birth, if clinically evident manifestations do not develop until later, the disease is usually not referred to as congenital. Hereditary, heritable, inherited, and genetic are essentially synonymous terms. In the past, the terms hereditary or inherited have been used in connection with dominant traits, that is, diseases in which the transmission from parent to progeny was directly evident; the term familial has been used for recessive traits which tend to affect members of single

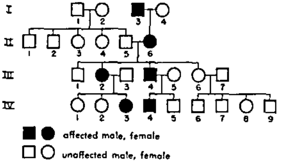


Fig. 11. Typical pedigree pattern of a rare autosomal dominant trait, classical achondroplastic dwarfism.

families, that is, sibs who are the offspring of unaffected parents. This terminology has little scientific rationale. A recessive trait is inherited as genuinely as is a dominant one.

Patterns of inheritance. The pedigree pattern characteristic of a rare autosomal dominant disease trait is shown in Fig. 11. Ideally, in such a case one-half the offspring of an affected individual are affected without regard to sex. An unaffected individual cannot transmit the trait. Most affected individuals are heterozygotes because homozygosity can result only from the mating of two affected persons, and this is unlikely for a rare trait. An example of dominant inheritance is provided by osteogenesis imperfecta, the syndrome of brittle bones, blue sclerae, and deafness.

The pedigree pattern characteristic of a rare autosomal recessive disease trait is shown in Fig. 12. The parents are unaffected, but each is a heterozygous carrier of the trait. Ideally, one-fourth of the offspring of such parents have the gene in homozygous state and are affected, regardless of sex; one-half the offspring, regardless of sex, are heterozygous carriers of the trait; one-fourth, regardless of sex, are homozygous normal. Among parents of sibships containing affected persons there is a higher incidence of consanguineous marriages, for example, first-cousin matings, than in the general population. Examples of recessive traits are albinism and alkaptonuria, both inborn errors of metabolism.

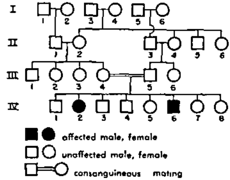


Fig. 12. Typical pedigree pattern of a rare autosomal recessive trait, one form of albinism.

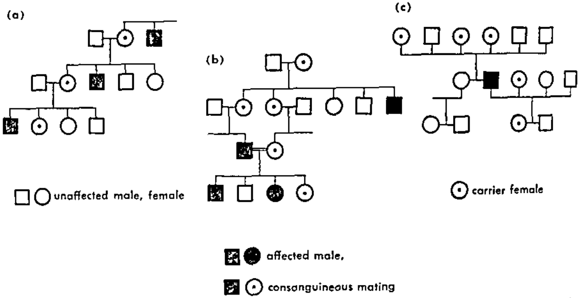


Fig. 13 Typical pedigree pattern of a sex-linked recessive trait, classical hemophilia. (a) Only males are affected and usually come from unaffected mothers who often had affected fathers. (b) An affected female can result from an affected father and a carrier mother. (c) All daughters of affected males are carriers.

The pedigree pattern characteristic of a rare sex-linked (X-linked) recessive trait is shown in Fig. 13. Usually only males are affected. Ideally one-half the sons of a carrier female are affected, and one-half the daughters of a carrier female are also carriers. The marriage of an affected male with a carrier female, which might happen if an affected male marries his first cousin, will result in one-half carrier females and one-half affected females. If an affected male marries a noncarrier female, all sons are normal and all daughters are carriers. The classic example is hemophilia. If, because of the gravity of the disease traits, affected males never reproduce (as in one form of muscular dystrophy), it may be impossible to be completely certain that the trait is not an autosomal sex-limited recessive.

Recessivity (recessiveness) and dominance are somewhat artificial distinctions. For many recessive diseases it is now possible by subtle tests to identify the heterozygous carriers. A true dominant, furthermore, should have the same expression whether the gene is present in either heterozygous or homozygous state. The gene in homozygous state has been observed for few dominant disease traits in the human species. For this reason some suggest using the designation conditional dominant or provisional dominant until the homozygote has been observed.

The arbitrary nature of dominance and recessivity is indicated by the situation with respect to sickle-cell trait, discussed previously. From the familial distribution of the hemolytic anemia without knowledge of the sickling phenomenon, it would seem to be a recessive. But from the distribution of the sickling phenomenon itself it would appear to be a dominant. Because of these considerations, sickling is sometimes referred to as an autosomal intermediate trait (Fig. 14).

Sex-influenced, or sex-limited, autosomal inheritance can be confused with sex-linked inheritance. A sex-influenced trait, although determined by a genetic factor located on an autosomal chromosome, behaves differently in men and in women because of hormonal or other physiological differences. Premature baldness behaves as a dominant in men but as a recessive in women; in men the heterozygotes show the trait, but in women only homozygotes show it (Fig. 15). Heberden's nodes, arthritic changes in the distal interphalangeal joints, develop as a dominant trait in postmenopausal women but in men the trait behaves as a recessive.

The pedigree pattern characteristic of a sex-linked (X-linked) dominant trait is illustrated in Fig. 16. Affected (heterozygous) females transmit the disease to half their offspring regardless of sex. On the other hand, affected hemizygous males have daughters all of whom are affected and no affected sons. The hemizygous male is likely to be more se-

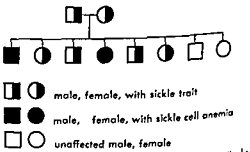


Fig. 14. Typical pedigree pattern in sickle-cell disease.

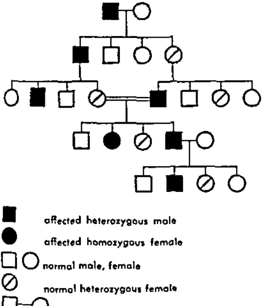


Fig. 15. Typical pedigree pattern of a sex-influenced autosomal trait, baldness.

verely affected than the heterozygous female. The presence of a "normal" allele on the second X chromosome of the female hence appears to have an alleviating effect. To call such a trait a sex-linked intermediate trait is, therefore, a better designation. Examples include hereditary hypophosphatemic rickets (a form of hereditary nephritis), choroideremia (an eye disease), and nephritis combined with deafness.

A factor located on the nonhomologous part of the Y chromosome would theoretically result in a disease which was transmitted to all males, and only to males, in each successive generation. No acceptable pedigree showing this so-called holandric (all-male) inheritance has been reported. Partial sex-linkage on the basis of genes located on the homologous segments of the X and Y chromosomes, although postulated for a number of traits, also has not been satisfactorily demonstrated in man.

What appears to be the same hereditary disease may be transmitted as an autosomal dominant in some families and as an autosomal recessive or sex-linked recessive in others. In genetic terms, one phenotype is not necessarily specific for one genotype. The phenotype is what the physician observes; the genotype is the actual genetic constitution responsible for the particular disease trait. The distinction between genotype and phenotype, one might say, is comparable to that between character and reputation. Often when such diseases are scrutinized clinically, at least slight phenotypic differences between the forms which are differently inherited become evident. Studies of genetic linkage provide another method for distinguishing genetic varieties of diseases which appear phenotypically identical. For example, elliptocytosis, a morphologic peculiarity of the red blood cell, is determined in some families by a gene at a locus on the same chromosome as the Rh locus; in other families there is no demonstrable linkage. Environmental insults also can result in a phenotype like that produced by a particular mutant gene. Such phenocopies, as they are called, can therefore be either genetic or environmental.

Genes and environment interaction. This is well demonstrated in many disease traits. Influencing the expression of mutant genes by manipulation of the environment is a leading method for treating genetic diseases.

The following are examples of inherited traits which never become evident unless the environment is appropriate. (1) An enzyme defect of the red blood cell renders it sensitive to hemolysis if the subject takes a drug, such as the antimalarial drug primaquine or the antibacterial agent Furadantin, or if he eats the bean of *Vicia faba*. (2) A genetic deficiency of serum pseudocholinesterase is no apparent inconvenience unless the subject is given the muscle relaxant succinylcholine in connection with a surgical operation; prolonged respiratory paralysis then results from failure to break down the administered agent at a normal rate.

In galactosemia the inability to metabolize adequately the galactose derived from milk results in dire effects such as mental retardation, cataract, and cirrhosis. One can imagine that if the customary diet of infants were free from milk the presence of the trait would never be recognized. In the genetic derangements of lipid metabolism, re- [illegible] in fat.

Variation in genetic diseases. Penetrance is an all-or-none concept related to whether an individual

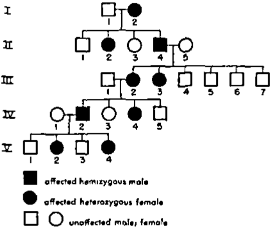


Fig 16. Typical pedigree pattern of a rare sex-linked dominant trait, hereditary hypophosphatemic rickets. Among the offspring of affected hemizygous males all daughters and no sons are affected.

with the appropriate genotype shows the disease trait. When an individual who from the evidence in his family tree (such as the presence of a dominant trait in one of his parents and in some of his children) does not himself show the trait, the gene is said to be nonpenetrant. A skipped generation is the result.

Expressivity refers to the clinician's "grade of severity." Especially in the instance of dominant traits, a wide range of severity (expressivity) may be observed in a group of patients. The largest number of patients will have disease of an intermediate severity but in some it may be very severe and in others very mild.

The interrelationship between penetrance and expressivity is illustrated in Fig. 17. Expressivity is indicated by a bell-shaped Gaussian curve. At the mild end of the curve there will be individuals so mildly affected that they will not, by the available methods of study, be recognizable as affected. The gene will be said to be nonpenetrant in such instances. Penetrance becomes more and more complete as the physician's methods for recognizing the presence of the gene become more refined.

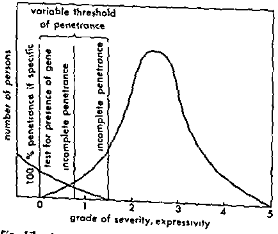


Fig. 17. Interrelationship of penetrance and expressivity.

The variability in the *expression* of a given gene is the result of the *rest of* the genetic make-up of the individual and of the *environment* in which the individual develops and *functions*. In some in- [illegible] result in different grades of severity. There can be not only different *alleles* of the gene primarily responsible *for the* derangement, but also, in the case of *dominant* genes, different alleles of the normal recessive member of the heterozygote pair.

Syndromes. Clinical genetics concerns itself with a considerable *number* of syndromes in which diverse manifestations occur *in the* same patient in a predictable manner. *The syndrome* of brittle bones, blue sclerae, and *deafness* has already been mentioned. Usually, if not always, these well-delineated syndromes are not the result of genetic linkage, that is, are not due to location of separate genes for each component close together on the same chromosome. Instead, such syndromes are virtually always the result of a single mutant gene in the heterozygous state (if dominant), homozygous state (if recessive), or hemizygous state (if sex-linked recessive). Because the gene controls some fundamental biochemical process of wide importance in the organism, diverse manifestations result. A strong argument for the unifactorial basis of syndromes (there is other evidence, as well) is provided in those conditions in which it is possible to relate all the clinical signs and symptoms to a single biochemical fault.

Sporadic cases. The clinician frequently observes isolated instances of a disease which in other families affects multiple members and is clearly inherited as a dominant. Several alternative possibilities must be considered in such so-called sporadic cases:

1. The case may represent a new dominant mutation.

2. The case may be a genetic phenocopy of the better known trait, and be inherited as an autosomal recessive in the particular family, with no sibs happening to be affected, or as a sex-linked recessive with no affected maternal uncles or other relatives. Even more complex genetic patterns are possible.

3. The case may be an environmental phenocopy.

4. One of the parents may in fact be mildly affected, but the methods for diagnosis are too insensitive to prove involvement. The gene is then, practically speaking, nonpenetrant in this parent.

5. The affected individual may be illegitimate. One of the putative parents (almost always the father) may not be the true parent.

6. A chromosomal aberration may be a *theoretically* possible basis, especially for very gravely affected sporadic cases.

Anticipation. Anticipation is a phenomenon which in the past was thought to have genuine biological significance. It was thought that dominant disease traits tend to express themselves earlier and in more grave form in each successive generation. It is now clear that anticipation is an artifact arising from the way in which families come under observation. With the rather wide variability of dominant traits, affected parent-child pairs will tend to be loaded with those cases in which the child has a severe form of the disease, with early onset, bringing him under medical observation. Those children with a mild disease which does not manifest itself until later in life are less likely to be represented *in the counts*.

Abiotrophy. Abiotrophy is a useful concept if one does not permit use of the term to lead to the impression that a mechanism has been elucidated. The term was invented by the neurologist W. R. Gowers to describe the fact that in certain hereditary disorders of the central nervous system, the

nervous system is functionally and probably anatomically normal early in life, but nonetheless has an inborn weakness which predisposes to deterioration along certain lines. An example is Huntington's chorea.

Chromosomal aberrations. A quantitatively abnormal chromosome complement has been identified in three disorders of man: mongoloid idiocy and two types of sex aberration.

In the mongoloid idiot 47 chromosomes are present rather than the usual 46. The extra chromosome is one of the smallest autosomal chromosomes. The unbalance produced by this so-called trisomic state is apparently the cause of this distressingly frequent disease. Nondisjunction during meiosis in the mother is thought to be the mechanism. Rather than a single chromosome of each pair going into the egg, the two homologous chromosomes of one particular pair stay together and both pass into the egg, which upon fertilization develops into the mongoloid individual. In *Drosophila* (the fruit fly) nondisjunction is known to occur more frequently in older organisms. In man it is known that the chance of bearing a mongoloid child is also much greater in older mothers. So far as is known, all monozygotic twins have been concordant for mongolism, whereas dizygotic twins are usually discordant. A few female mongoloid idiots have had children; the child has also been affected in several reported instances.

The possibility that other congenital malformations and disorders of man may be caused by a chromosomal aberration is under active investigation. Qualitative abnormalities of the chromosomes, such as deletions and translocations, will undoubtedly be detected as technical methods improve and experience increases.

The two sex aberrations in which abnormality of chromosome count has been identified are the Turner syndrome, in which there is only one sex chromosome, an X, the individuals are said to be XO, and the total chromosome count is 45; and the Klinefelter syndrome, in which there are three sex chromosomes, XXY, and the total count is 47. These conditions are of great theoretical interest because they indicate that sex determination in man is different from that in *Drosophila*. In man (and the same is true for the mouse and probably other mammals) the Y chromosome determines male phenotype, or at least the development of masculine external genitalia. [V.A.M.]

Bibliography: B. Childs and J. B. Sidbury, Jr., A survey of genetics as it applies to problems in medicine, *Pediatrics*, 20:177–218, 1957; V. A. McKusick, *Heritable Disorders of Connective Tis-* [illegible] *Heredity*, 1954; R. R. Race and R. Sanger, *Blood Groups in Man*, 3d ed., 1958; A. G. Steinberg, Methodology in human genetics, *J. Med. Educ.*, 1958; C. Stern, *Principles of Human Genetics*, 1949.

Humidification

The process of increasing the water-vapor content (humidity) of a gas. This process and its reverse operation, dehumidification, are important steps in air conditioning for human comfort and in many industrial operations. For a discussion of the dehumidification of gases other than air, *see* DRYING.

Humidity is normally defined as the weight of water (in pounds or grains) carried by 1 lb of dry air (sometimes called absolute humidity). The water content of air also is frequently indicated as a relative humidity, which is the ratio (expressed as a percentage) of the partial pressure of water vapor in the air to the vapor pressure of pure water at the same temperature.

Air (or other gas) can be humidified by direct injection of water vapor (steam) or, more commonly, by the evaporation of liquid water in contact with the air stream. When evaporation occurs, heat is required to provide the latent heat of vaporization. If no external source of heat is provided, either the water or the air, or both, will be cooled. The cooling of water by this process is the basis of operation for industrial cooling towers, whereas evaporative air coolers often used in hot, dry climates depend upon the air-cooling effect. In both these types of apparatus, humidification of the air occurs although it is not the prime objective of the operation. In units designed primarily for humidification, the incoming air is usually heated to provide the latent heat of evaporation and to permit the air to leave the unit at both controlled temperature and humidity.

Psychrometric chart. The operation of humidification equipment can best be understood by reference to a psychrometric chart as shown in Fig. 1 (*see* PSYCHROMETRICS). Every point on this chart represents a specific air condition with regard to temperature (bottom scale) and humidity (right-hand scale). Several other characteristics of the air are indicated by the location of the points on the chart, including the relative humidity and the wet-bulb temperature. The latter is defined as the equilibrium temperature attained by a small surface of liquid evaporating into a large amount of unsaturated air. It is normally measured by moving air rapidly past a thermometer bulb covered with a wetted wick—hence the name. For the air-water system, the wet-bulb lines can also be used with sufficient accuracy to indicate the adiabatic saturation temperature, the temperature which a gas will attain when saturated by adiabatic contact with water which is already at the adiabatic saturation temperature.

Operation of an air conditioning humidifier such as shown in Fig. 2 can be represented by the path A-B-C-D on the psychrometric chart. Air entering at the conditions indicated by point A is first heated to point B without change in humidity by contact with a heated coil. It then passes through a water-spray zone and is adiabatically cooled and humidified to the conditions represented by point C. T

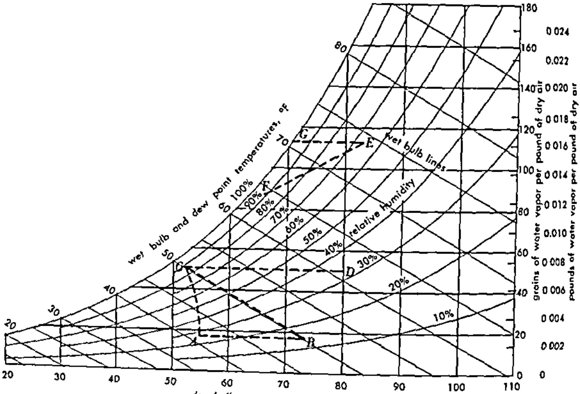


Fig. 1. Psychrometric chart.

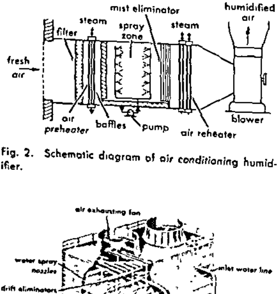


Fig. 2. Schematic diagram of air conditioning humidifier.

Fig. 3. Cutaway view of large industrial counterflow cooling tower. (Courtesy Fluor Products Co.)

humidified air is then heated to the desired final outlet conditions of point *D*. Alternatively, the entering air can be contacted with heated water so that humidification occurs with little or no cooling. This path is represented by line *A-C*.

Cooling towers. Line *A-C* also represents a typical cooling tower operation in which warm water is contacted with air for the purpose of cooling the water. Cooling towers are of considerable industrial importance in power plants, refineries, chemical plants, and large air conditioning or refrigeration installations in which considerable quantities of water are used in condensers and coolers to remove process heat. Two general types of water cooling towers are employed—natural circulation and mechanical draft. The former depend primarily upon wind and natural draft effects to provide air circulation, whereas the latter employ fans to move air through the tower. Cooling towers may be further subdivided into crossflow designs, in which the air moves horizontally while the water falls vertically, and counterflow designs, in which the air moves upward countercurrent to the falling water. The crossflow design is useful for towers which must be kept to a minimum height; however, counterflow operation is theoretically more efficient, and this type of tower is capable of producing colder water. A cutaway illustration of a commercial counterflow cooling tower is shown in Fig. 3. See COOLING TOWER.

Dehumidification. The dehumidification of air is indicated by line *E-F* on the psychrometric chart.

Dehumidification may be accomplished by contacting the air with cold water in a device similar to that of Fig. 2, except that the heater coils are not necessary, or by passing the air across banks of finned tubes through which cold water or refrigerant is passed. Because water condenses on the outside walls of the tubes, the two processes are equivalent in that the warm, moist air is in direct contact with cold water. Condensation results in the release of the latent heat of condensation and this raises the temperature of the cooling liquid. It is not necessary for the entire gas stream to be cooled to the dew point (point G) for condensation to occur; however, no water can be condensed from the gas unless the cooling surface is below the dew point.

Dehumidification can also be accomplished by the use of solid desiccants such as silica gel or alumina, or liquid absorbents such as triethylene glycol or lithium chloride solutions. Both types are commonly employed in regenerative systems in which the spent dehydration agent is stripped of water by the application of heat. *See* AIR CONDITIONING; COMFORT CONTROL; DESICCANT; GAS ABSORPTION OPERATIONS; HEAT EXCHANGER; REFRIGERATION; STRIPPING; UNIT OPERATIONS.

[A.L.K.]

Bibliography: John H. Perry (ed.), *Chemical Engineers' Handbook*, 3d ed., 1950.

Humidistat

A controller that measures and controls relative humidity. A humidistat may be used to control either humidifying or dehumidifying equipment by the regulation of electric or pneumatic switches, valves, or dampers

Most methods for measuring humidity rely upon the swelling and shrinking of materials, such as human hair, silk, horn, goldbeater's skin, and wood, with increases and decreases in relative humidity.

Human hair is most commonly used because of its small diameter, which contributes to rapid absorption and dissipation of moisture. Strands of hair are bunched and several such bunches are combined in a ribbonlike element (Fig. 1).

As the relative humidity of the air decreases, the strands of hair shorten; this movement is transmitted through a suitable lever mechanism to an electric switch or pneumatic valve, which is part of the humidistat.

An electronic humidistat includes a sensing element and a relay amplifier. The sensing element consists of alternate metal conductors on a small, flat plate with a plastic coating (Fig. 2).

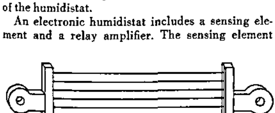


Fig. 1. Hair element humidistat. *(Minneapolis-Honeywell Regulator Co.)*

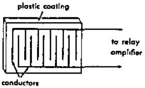


Fig 2. Electronic humidistat. *(Minneapolis Honeywell Regulator Co.)*

An increase or decrease of the relative humidity causes a decrease or increase in the electrical resistance between the two sets of conductors; the change in resistance is measured by the relay amplifier. Small changes in relative humidity can be measured in this way for precise control.

[J.E.HA.]

Bibliography: J. E. Haines, *Automatic Control of Heating and Air Conditioning*, 1953.

Humidity

Atmospheric water-vapor content, expressed in any of several measures, especially relative humidity, absolute humidity, humidity mixing ratio, and specific humidity. Quantity of water vapor is also specified indirectly by dew point (or frost point), vapor pressure, and a combination of wet-bulb and dry-bulb (actual) temperatures

Relative humidity is the ratio, in per cent, of the moisture actually in the air to the moisture it would hold if it were saturated at the same temperature and pressure. It is a useful index of dryness or dampness for determining evaporation, or absorption of moisture. *See* PSYCHROMETRICS.

Human comfort is dependent on relative humidity on warm days, which are oppressive if the relative humidity is high but may be tolerable if it is low. At other than high temperatures, human comfort is only slightly affected by variations in relative humidity. *See* TEMPERATURE-HUMIDITY INDEX.

Absolute humidity is the weight of water vapor in a unit volume of air expressed, for example, as grams per cubic meter or grains per cubic foot.

Humidity mixing ratio is the weight of water vapor mixed with unit mass of dry air, usually expressed as grams per kilogram. Specific humidity is the weight per unit mass of moist air and has nearly the same values as mixing ratio.

Dew point is the temperature at which air becomes saturated if cooled without addition of moisture or change of pressure; frost point is similar but with respect to saturation over ice. Vapor pressure is the partial pressure of water vapor in the air. Wet-bulb temperature is the lowest temperature obtainable by whirling or ventilating a thermometer whose bulb is covered with wet cloth. From readings of a psychrometer, an instrument composed of wet- and dry-bulb thermometers and a fan or other means of ventilation, values of all other measures of humidity may be determined

from tables. *See* DEW POINT; EVAPORATION; FOG; PRECIPITATION (METEOROLOGY); VAPOR PRESSURE. [J.R.F.]

Bibliography: R. J. List (ed.), *Smithsonian Meterological Tables*, 6th rev. ed., 1951.

Humidity control

The addition or removal of water vapor associated with a quantity of dry air. Change of absolute humidity (the weight of water vapor present per pound of dry air) necessarily changes the enthalpy of the air-vapor mixture and hence is a heating process, a cooling process, or an adiabatic (no external heat exchange) process (*see* AIR COOLING).

Humidification. Simple humidification (increase of latent heat without change in sensible heat) can be accomplished by passing the air-vapor over a wetted coil (or through a high-rate stream of wash water) when the coil surface (or droplet surface) is maintained at the dry-bulb temperature of the air. Other simple methods of humidification involve the discharge directly into the air-vapor stream of the desired quantity of water in the form of steam at a calculated balance state (a state in which, at equilibrium of admitted steam and passing air-vapor, the atmosphere will be at an unchanged dry-bulb temperature), or, in domestic heaters, the direct evaporation of water from an open pan placed in a hot air furnace or from trays placed on radiators.

Dehumidification. Two principal methods for decreasing the water vapor associated with a quantity of dry air are absorption and adsorption.

Absorption is a means of dehumidification in which a moist air stream passes through a spray which utilizes a liquid sorbent that undergoes physical or chemical change as it removes water vapor from the atmosphere; typical liquid sorbents are lithium chloride solution, calcium chloride solution, and various ethylene glycols.

Adsorption is a means of dehumidification in which a sorbent such as silica gel or activated bauxite undergoes neither chemical nor physical change, but tends, often through capillary action, to reduce the surface vapor pressure and lead to condensation of water vapor from the air stream within its capillary openings. *See* ADSORPTION; DEHUMIDIFIER. [F.W.HU.]

Humite

A series of magnesium nesosilicate minerals closely related in crystal structure and chemical composition. The group comprises four minerals as shown in the accompanying table.

Composition and cell dimensions of minerals in humite group

Mineral	Composition	a_0	b_0	c_0
Norbergite	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Norbergite and humite are orthorhombic; chondrodite and clinohumite are monoclinic; all have closely related axial lengths. As shown in the table the cell dimensions are nearly the same in the b and c directions, while the a dimension is related to the number of magnesium atoms in the formula. *See* SILICATE MINERALS.

The minerals of the humite group have similar physical properties. The luster is resinous and the color usually light yellow to brown, more rarely white or red. Hardness is 6–6½ on Mohs scale; specific gravity is 3.1–3.2. Since the several minerals occur under similar conditions, it is difficult to distinguish them by inspection. They are found characteristically in contact zones in limestone. Norbergite, found at Norberg, Sweden, is the rarest of the group. The others have all been found in the ejected material at Mount Vesuvius. Fine crystals of chondrodite have come from the Tilly Foster iron mine at Brewster, New York. [C.S.HU.]

Hummingbird

Any of over 300 species of the American family Trochilidae, represented in the United States by 15 species. Most hummingbirds are tropical, small, long-billed, and brilliantly colored. They feed on insects and nectar, hovering before flowers as they

The ruby-throated hummingbird, *Archilochus colubris*, length to 4 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

feed. Of the United States species, all are western except the ruby-throated hummingbird, *Archilochus colubris*. Like many others, it is iridescent green above, and the male is characterized by an iridescent ruby-red throat patch. Other species show a fascinating array of brilliant colors. *See* APODIFORMES. [J.D.B.]

Humus

The amorphous, ordinarily dark-colored, colloidal matter in soil representing a complex of the fractions of organic matter of plant, animal, and microbial origin that are most resistant to decomposition.

Humus consists of the combined residues of or- [illegible] original struc- [illegible] sim- [illegible] sub-stance as well as by-products of microorganisms

It is not a definite substance and is in a continual state of flux, disappearing by slow decomposition, and being constantly renewed by incorporation of further residual matter. With a balance between these processes, humus, though not static, remains relatively uniform in nature and amount in a given soil. It constitutes a reservoir of stabilizing material which imparts beneficial physical, chemical, and biological properties to soil. Fertile soils are rich in humus.

Humus improves the texture of soils. It exerts a binding effect on sandy soils, and loosens the harder, clayey soils, thus increasing their porosity and permeability. It increases the moisture-holding capacity and improves the granular structure by cementing mineral particles into stable crumbs. This helps soils resist the pulverizing and eroding action of wind, water, and cultivation. As a storehouse of elements important to plants, humus functions as a regulator of soil processes by liberating gradually nutrients that would otherwise drain away. A soil rich in humus provides optimum conditions for the development of beneficial microorganisms and constitutes the best medium for growth of plants.

Humus formation depends upon an adequate supply of raw organic residues and upon suitable conditions for their decomposition. The composition of humus varies with the nature of the surface vegetation which furnishes the bulk of its raw supply. The degree of decomposition of residues, and the nature of the humus formed, depend also on soil and climatic factors, chief of which are temperature, aeration, and moisture. Under optimum conditions bacteria, actinomycetes, and fungi engage in the rapid breakdown of primary organic matter. With excessive moisture and poor aeration the aerobic microorganisms are relatively suppressed and decomposition is only partial.

Peat. This is a type of humus that results from the decomposition of plant material under conditions of excessive moisture or in areas submerged in water. It is an organic deposit formed in marshes and swamps by the partial decomposition of countless generations of a variety of plants. The water, by excluding air, prevents rapid oxidation, and decay results from the action of more restricted groups of microbes than in mineral soils. Anaerobic bacteria are prominent, effecting partial mineralization with much liberation of gases. In the course of time large deposits are built up, often in layers as the nature of the marsh vegetation alters. This sedimentary lowmoor peat, slightly acid, may by proper drainage be turned into good agricultural soil. Highmoor peat, formed in circumstances providing for less sedimentation, largely from moss-type plants, is more fibrous, poorer in nitrogen and minerals, and very acid. The microbial content is low and conversion to agricultural use more difficult.

Organic decomposition and humification. When organic matter such as plant stubble, green manures, farmyard manures, composts, or organic fertilizers is incorporated in soil, microorganisms begin to decompose it. Some ingredients are readily attacked while others are very resistant. Sugars, starches, organic acids, and alcohols are most rapidly destroyed, followed by fats, cellulose, and hemicelluloses as well as the less-resistant proteins. More resistant to microbial attack are waxes, certain proteins, and lignins, particularly. Lignins constitute increasingly higher proportions of the organic residues as humification proceeds. During decomposition there is considerable synthesis of microbial cell substance as microorganisms increase. As certain groups are superseded by others with the changing nature of the residual matter, large numbers of dead cells are incorporated into the organic complex. This has the effect of increasing the protein content. See CELL WALLS IN PLANTS.

During the course of decomposition the carbon-nitrogen ratio, which may vary from 12:1 to 20:1 in green manure crops, to as much as 80:1 in straw, gradually narrows with the liberation of carbon dioxide until a level of approximately 10:1 is reached which, in most mineral soils, is characteristic of humus.

Carbohydrates. These compounds are readily attacked by a wide variety of bacteria, actinomycetes, and fungi in the presence of sufficient oxygen. The disaccharides and starches are first hydrolyzed with formation of monosaccharides which are oxidized, and the carbon is finally liberated as carbon dioxide. With limited air supply anaerobic bacteria are chiefly responsible for decomposition. In this case the rate is slower, the chemical changes less complete, and in addition to carbon dioxide, less completely oxidized products such as methane, organic acids, and alcohols are formed. Under favorable conditions these may in turn be completely oxidized by aerobic microorganisms. See CARBOHYDRATE.

Cellulose. This is a polysaccharide found in the fibrous structure of many plants. Cellulose is decomposed chiefly by bacteria and fungi in a manner similar to that described previously for carbohydrates. However, it is resistant to attack by the majority of soil microorganisms, the types capable of producing cellulose-hydrolyzing enzymes being restricted. Cellulose decomposes much more slowly than sugars or starches but the end products of aerobic and anaerobic decomposition respectively are much the same. During decomposition of non-nitrogenous organic compounds, such as cellulose, the organisms involved demand adequate nitrogen. Consequently the addition of excess amounts of cellulosic material to soil may result in temporary loss of mineral nitrogen to plants through its being assimilated by microbes into cell substance. See CELLULOSE.

Hemicelluloses. These are important constituents of plant material but unlike cellulose, they vary in chemical composition. They consist of polysaccharides combined with sugar acids of the uronic ty[illegible] and are related to the pectins. They are m[illegible] ily attacked by microorganisms than cell[illegible]

the rate of decomposition varies with their chemical nature. On hydrolysis they yield monosaccharides in the form of hexose and pentose sugars and organic acids and alcohols. These products are readily decomposed further. *See* HEMICELLULOSE.

Proteins. These compounds are decomposed by large numbers of soil microorganisms, comprising fungi, actinomycetes, and aerobic and anaerobic bacteria. On hydrolysis by microbial enzymes, proteins are split first into polypeptides and then into simpler amino acids. These are readily attacked further with the result that ammonia appears as the chief nitrogenous by-product of respiration. For this reason, the process of protein decomposition is referred to as ammonification. This step is a necessary preliminary to nitrification, which results in the oxidation of ammonia to nitrate. Under the aerobic conditions existing in most arable soils the products of oxidation include also carbon dioxide and sulfates. Under anaerobic conditions, where oxidation is incomplete, decomposition products include hydrogen sulfide, amines, mercaptans, and such compounds as indole and skatole, as well as organic acids and alcohols. *See* NITROGEN CYCLE; PROTEIN.

Lignin. A polymerized compound of high molecular weight composed of carbon, hydrogen, and oxygen, lignin is of uncertain chemical composition. It represents the plant ingredient most resistant to decomposition. It is an important constituent of straw and wood, forming 5–30% of the tissue, with the percentage rising as the plant matures. Few microorganisms are able to attack lignin and these are chiefly higher fungi. Hydrolysis and oxidation of lignin are slow and the by-products of its degradation are not well known. As plant material decomposes lignin accumulates to constitute an important ingredient of humus, though in a modified condition. In the process of humification lignin unites with proteins, chiefly those elaborated by microorganisms, to form a lignin-proteinate complex. This renders the protein very resistant to microbial attack. Thus nitrogen in humus is retained and only slowly liberated.

Humus itself undergoes slow decomposition so that in well-aerated soils receiving normal additions of organic material its level remains fairly constant. This slow decomposition is of great practical importance. It constitutes a means whereby nitrogen, as well as simple mineral compounds, can be stored in soil and released gradually for consumption by growing plants. *See* SOIL MICROBIOLOGY.

[A.C.L.]

Hunger

Although the term hunger is most commonly used to refer to the subjective *feelings* that accompany the need for food, the study of this topic has come to include consideration of the over-all regulation of food intake. More specifically, experimental work on the problem of hunger has been concerned with (1) the sensory cues that give rise to feelings of hunger, (2) the physiological mechanism that determines when and how much food will be ingested, and (3) the mechanism governing the selection of the type of foodstuff to be ingested.

Cannon theory. The earliest experimental approach to these problems concentrated almost exclusively on the question of the sensations of hunger or, as they have come to be known, the hunger pangs. This early work by W. B. Cannon, which resulted in his so-called local theory, led to the conclusion that for both hunger and thirst the appropriate sensations arose peripherally in the body. According to this theory, the hunger pangs were stomach contractions that produced the sensations of hunger through stimulation of the local sensory nerves. More recent work substantiates the fact that increased stomach contractions do indeed accompany the state of hunger in ordinary circumstances, but it seems unlikely that these contractions contribute substantially to the detailed regulation of eating behavior. For example, when the sensory nerves of the stomach are cut or even when the stomach is altogether absent, eating behavior can go on in an essentially normal manner. Whatever influence stomach contractions may have on the ingestion of food, it is known that the motility of this organ can be controlled by both a neural and a hormonal route. The hormone involved may be one that is secreted by the stomach itself. *See* HORMONE; NERVOUS SYSTEM; THIRST.

Physiological mechanism. It is clear that food consumption is most basically regulated by the nutritional status of the organism. Food deprivation leads to eating, and the ingestion of food materials terminates the state of hunger. The question is to determine what physiological process it is that varies quantitatively with the nutritional status of the individual and is capable of influencing the nervous system in a manner that would instigate and terminate food consumption at the right times. The results of attempts to find a simple humoral factor that might be involved linearly in this regulation have not been illuminating.

Blood sugar level. Blood sugar level, which has received more attention than any other factor, can be used as a case in point. The concentration of blood sugar does indeed vary appropriately in a general way with the periodicity of the food cycle. Moreover, hyperglycemia, or extremely high blood sugar, and hypoglycemia, or extremely low blood sugar, have been observed to decrease and increase hunger respectively. The detailed analysis of the normal life variations of blood sugar, however, reveals that the relation between the concentration of blood sugar and hunger is not sufficiently close for this single, humoral factor to be able to control hunger in any simple and direct manner. This same point could be made also about the other humoral factors that have been investigated.

Tissue utilization of food. The evaluation of tissue utilization of food materials seems a more promising approach to this problem. Thus far only glucose has been studied. When tissue utilization of glucose is small and there are, accordingly, large

reserves of sugar in the cells, the individual is reported to feel satiated as far as food is concerned. Conversely, when the cells increase their utilization of sugar and the cellular reserves start to be depleted, feelings of hunger arise. Tissue utilization of glucose in such experiments has only been inferred by comparing the blood sugar level in arterial and venous blood samples. What is known about the neural mechanism involved in hunger and eating behavior adds further support to the belief that body glucose is involved in some manner. It is not clear whether other nutrient materials such as, for example, proteins and fats have their own comparable mechanisms through which they can influence the level of hunger or whether they affect hunger only through their known ability to modify the availability of body glucose.

Neural centers. There is considerable information concerning the neural centers involved in regulating food consumption. There are at least two central nervous system (CNS) *nuclei*, in the hypothalamus, that are clearly involved. The ventromedial nucleus appears to be a satiation center, while a cluster of cells located at the same level of the hypothalamus but situated more laterally constitutes a so-called feeding center. Bilateral surgical damage to the ventromedial nucleus results in hyperphagia or overeating. Rats having such experimental lesions pass through three successive stages: (1) For the first day postoperatively, the animals attack available food materials ravenously and eat large quantities (2) For the subsequent month or two, the animals settle down to persistent eating and increase to several times their normal weight. This period is called the dynamic phase. (3) Finally, during the static phase, the rats return to the consumption of normal quantities of food but keep the level of obesity they had attained during earlier stages.

Conversely, bilateral lesions in the feeding centers produce anorexia or absence of appetite. In some experimental animals of this type, food is not swallowed even though it is placed directly in the mouth. As would be anticipated, electrical stimulation of this center by way of electrodes buried in the brain results in increased eating. The hyperphagia produced in this manner persists a number of hours beyond the actual period of stimulation. When both the satiation centers and the feeding centers are removed, anorexia is still the result. This suggests that the satiation center normally operates by inhibiting the feeding center. There is also an area more anterior in the hypothalamus from which stomach contractions can be elicited by electrical stimulation. Further, lesions in the anterior midbrain have been reported to produce hyperphagia. Whether this latter area is functionally related to the ventromedial nucleus is not known. *See* BRAIN.

There is reason to believe that the ventromedial nucleus of the hypothalamus has its activity influenced by the concentration of its intracellular glucose. The suggestion is that when there is ample glucose within the cells this satiation center has its electrical activity accordingly increased and in some manner produces satiation in the organism. This theory is dramatically supported by the fact that the intravenous injection of gold thioglucose (a compound formed by linking glucose with a chemical having known toxic effects on body cells) results in specific damage to the ventromedial nucleus. Further, when amphetamine, which is known to reduce the appetite for food, is injected into the body, it specifically produces increased electrical activity in the ventromedial nucleus. It is possible that nutrients other than glucose may also be capable of influencing these or other related neural centers but this has not been established.

Specific hunger behavior. Deprivation of certain, specific food substances results in an increased appetite for the particular substance which the organism needs. This indicates that in some manner it is possible for many substances other than glucose to influence the regulation of eating behavior. This so-called specific hunger behavior has been demonstrated experimentally in connection with many substances such as, for example, salt, calcium, fats, proteins, and certain vitamins. It appears that in adult humans learned tastes have considerably masked the ability to be influenced by specific food needs, but this ability has been clearly demonstrated in children and in all lower animals studied. Theories to account for regulation of eating behavior by specific food deprivations have suggested the possibilities of learning, differentially lowered taste thresholds, and direct effects upon the brain.

Under at least one condition in which a specific food preference has been carefully studied, it is clear that neither learning nor changed taste thresholds are involved. Experimental animals surgically deprived of their adrenal glands spontaneously ingest large amounts of salt ($NaCl$), and the extra sodium (Na) thus obtained is essential to their survival without the adrenals. It seems that either salt tastes better to the operated animals or else the increased consumption is due to some effect of the salt after it is ingested. Closely related to this phenomenon of specific hunger behavior is the fact that when normal animals are presented with different concentrations of some nutrient solution, such as glucose or salt in water, they ingest more of certain concentrations than others and even avoid highly concentrated solutions. The quantity of these various concentrations that the animal will ingest under standard conditions is taken to represent the preference-aversion function for the solution under study. In such circumstances the volumes ingested are, in part, governed by the colligative characteristics of the solutions rather than by any specific chemical influences of the solute itself. The taste of the solution also affects intake when more than one solution at a time is available to the organism. *See* BODY RHYTHM. [R.A.M.]

Bibliography: C. T. Morgan and E. Stellar, *Physiological Psychology*, 2d ed., 1950.

Hunting

An instability phenomenon often observed in feedback control systems. The values of various system quantities oscillate over a sufficient period of time and with sufficient amplitude as to be considered undesirable. Hunting arises from a number of causes, such as friction, backlash, dead space, and incorrect gain setting. It can usually be overcome by proper over-all system design, manufacture, and adjustment. *See* CONTROL SYSTEMS. [J.A.H.]

Huntington's chorea

A rare hereditary disease of the basal ganglia and cerebral cortex, resulting in choreiform (dance-like) movements, intellectual deterioration, and psychosis. *See* HUMAN GENETICS.

The incidence is approximately 0.1% of hospital admissions. The illness is strictly hereditary, as a dominant Mendelian trait. All American patients are supposed to be descendants of two brothers with Huntington's chorea, who migrated to the United States from England.

The signs are choreiform movements, varying from restlessness to severe grimacing and incessant grotesque motion. The motions disappear in sleep. Patients develop a progressive intellectual impairment and paranoid delusions. Most patients require hospitalization in psychiatric institutions (*see* DELUSION; PARANOID STATE).

The diagnosis is clear-cut; differential diagnosis from other forms of chorea is usually simple. No effective treatment is known. *See* PSYCHOSIS.

[F.C.R.]

Bibliography: A. N. Bruce (ed.), *Neurology*, vol. 2, 2d ed., 1955.

Huntite

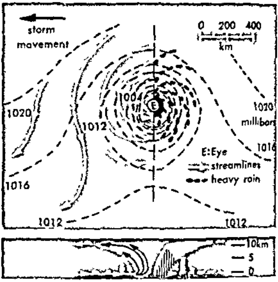
A very rare magnesium-calcium carbonate, $Mg_3Ca(CO_3)_4$, of low-temperature formation. Huntite has been reported from Nevada, Hungary, and Australia. It occurs in fine-grained masses with a distinct x-ray powder-diffraction pattern. *See* CARBONATE MINERALS. [R.I.H.]

Hurricane

A tropical cyclone of great intensity; called a typhoon in the tropical North Pacific. Any wind reaching a speed of more than 75 mph is said to have hurricane force. Many tropical cyclones do not reach that intensity.

Winds up to 160 mph have been measured in hurricanes. Enormous destruction results from both winds and tidal surges in coastal areas, and to a lesser extent from effects of heavy rains. Loss of life, which has been huge in many storms, has been cut down markedly by increasingly adequate warnings, based largely on increased upper-air observations and more accurate location by radar and aircraft reconnaissance.

Formation and movement. At the earth's surface, the hurricane appears as a nearly circular but somewhat asymmetrical vortex, with winds spiraling in toward the low-pressure center (see illustration) with a counterclockwise rotation (in the Northern Hemisphere). Both wind force and pressure gradient increase rapidly toward the center. Radius of the cyclonic circulation is generally less than 500 km, hurricane-force winds in a mature storm being confined to a ring inside a 50- to 150-km radius. Pacific typhoons are sometimes very large, with circulations extending to 1000 km from the center. In the eye of a tropical storm, averaging 25 km across, winds are weak, sometimes nearly calm.



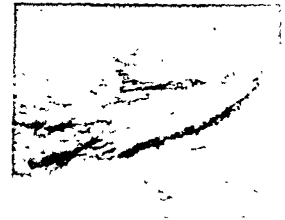
Simplified model diagram of a Northern Hemisphere hurricane circulation. Isobars (dashed lines) have been omitted near cyclone center. On left, streamlines of air flow; on right, regions of heavy rain. Below is a vertical section through center showing clouds and vertical circulation.

Central pressures of 950–960 millibars are not unusual, and pressures as low as 890 millibars have been recorded. Rainfall of 10–20 in. is common near the path of hurricane centers.

Storm tides or surges, resulting from stresses between wind and waves, are most pronounced on the right-hand forward side of a hurricane (with respect to its direction of motion) and may amount to 10–20 ft in coastal waters. *See* STORM SURGE.

Almost all hurricanes form between latitudes 5–25° over warm tropical seas. Formation is most frequent in the doldrums or intertropical convergence, a region of weak winds between trade-wind systems of the Northern and Southern Hemispheres, when the doldrums are farthest north or south of the equator in late summer or fall. No hurricanes are observed in the South Atlantic or the eastern South Pacific.

While in the tropics, hurricanes are located on the equatorward side of the subtropical anticyclones (*see* STORM), where they are imbedded in large-scale easterly currents and drift slowly (10–15 mph) westward. On the western sides of the oceans, most recurve poleward, finally moving rap-



View from airplane looking over the turbulent cloud patterns at the top of Hurricane Gracie in September, 1959. (*U.S. Navy photograph*)

idly eastward after entering the middle-latitude west-wind belt. In many cases, these regenerate into strong extratropical cyclones after drawing polar fronts into their circulations. Tracks of individual hurricanes are highly variable and often erratic. Only a minority of the total number move inland over continents, and in most cases these dissipate or weaken rapidly.

Structure and cloud systems. The hurricane, unlike an extratropical cyclone, is a warm-core storm. Air converging toward the center in low levels rises and diverges outward in upper levels (as shown by the section of the illustration) in a direct thermal circulation which produces energy to drive the storm. Warmth of the central portion (outside the eye) depends upon high water-vapor content in the rising air, where latent heat is given up on condensation.

In low levels, air converges toward the cyclone center in response to the strong centrally directed horizontal pressure-gradient force. Because pressure decreases more rapidly with height in cold dense air than in warm, the pressure gradient becomes weaker with increasing elevation. In converging in low levels, the air acquires cyclonic circulation and high angular momentum as a consequence of the Coriolis acceleration. Angular momentum is to some extent conserved when the air rises near the core, and in upper levels outward acceleration takes place because the centrifugal force exceeds the weak pressure-gradient forces there. This upper-level removal of air is essential to maintain the low pressure in the cyclone, which otherwise would fill up because of influx of air in low levels.

The main cloud system consists of a ring of heavy nimbostratus extending 30,000–40,000 ft, with heavy precipitation in the band of strong winds surrounding the eye. The base of the main cloud deck rises gradually away from the center, and the clouds become decreasingly thick, with altostratus and finally cirrostratus on the fringes. Broken cumulus and fractocumulus are found in low levels. Interspersed with these cloud systems are lines of tall cumulonimbus with heavy rain (lower part of section in illustration) spiraling inward toward the center.

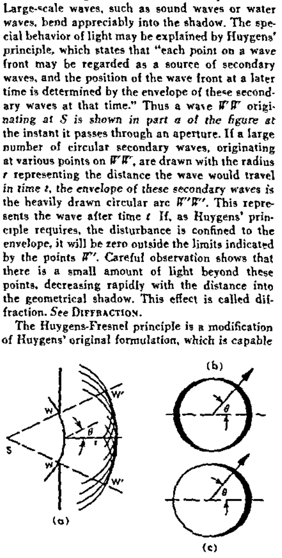
In a conical core inside the main cloud deck, there is an eye, or clear area in upper levels, where the air is warm and dry. Within the eye, dense, broken low clouds are found, but these do not extend to great heights. [C.W.N.]

Bibliography: T. F. Malone (ed.), *Compendium of Meteorology*, 1951; H. Riehl, *Tropical Meteorology*, 1954; I. R. Tannehill, *Hurricanes*, 9th ed., 1956.

Huygens' principle

An assumption regarding the behavior of light waves, originally proposed by C. Huygens in the seventeenth century to explain the fact that light travels in straight lines and casts sharp shadows. Large-scale waves, such as sound waves or water waves, bend appreciably into the shadow. The special behavior of light may be explained by Huygens' principle, which states that "each point on a wave front may be regarded as a source of secondary waves, and the position of the wave front at a later time is determined by the envelope of these secondary waves at that time." Thus a wave WW originating at S is shown in part *a* of the figure at the instant it passes through an aperture. If a large number of circular secondary waves, originating at various points on WW, are drawn with the radius r representing the distance the wave would travel in time t, the envelope of these secondary waves is the heavily drawn circular arc $W'W'$. This represents the wave after time t. If, as Huygens' principle requires, the disturbance is confined to the envelope, it will be zero outside the limits indicated by the points W'. Careful observation shows that there is a small amount of light beyond these points, decreasing rapidly with the distance into the geometrical shadow. This effect is called diffraction. See DIFFRACTION.

The Huygens-Fresnel principle is a modification of Huygens' original formulation, which is capable



Huygens' principle. (a) The construction for a spherical wave. (b,c) Amplitude of the secondary wave according to Fresnel and Kirchhoff, respectively.

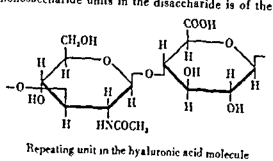
of explaining diffraction. A. Fresnel postulated that the amplitude of any secondary wave decreases in proportion to cos θ, where θ is the angle between the normal to the original wavefront and any point on the secondary wave (see figure, part *b*, where the thickness of the arc indicates the amplitude). Fresnel then modified Huygens' requirement that the disturbance be confined to the envelope, by specifying that at any point the disturbance was the resultant of all displacements due to secondary waves reaching that point. In this way Fresnel was able to explain the complicated diffraction patterns that are produced by sending light through small apertures. Subsequent theoretical investigations by G. Kirchhoff showed that the correct obliquity factor should be $1 + \cos\theta$ instead of cos θ (see figure, part *c*). Approximations made by Fresnel had compensated for this error. A discrepancy in the phase of the resultant wave, amounting to one-quarter period, was also explained by Kirchhoff's treatment. [F.A.J.]

Hyades

A group of stars known from early times, scattered through the V in the constellation of Taurus. An estimated 350 stars form the Hyades cluster at a distance of 130 light years from the Sun. Half the mass of the cluster is in a sphere of diameter 40 light years. The brightest stars of this galactic cluster are yellow to red. It is also the most prominent moving cluster. Known as the Taurus moving cluster, the group has a velocity of 45 km/sec toward a convergent point in the constellation of Orion. Strong similarity between the Hyades and the galactic cluster Praesepe leads to hypothesis of common origin. [H.S.H.]

Hyaluronic acid

This polysaccharide is an integral part of the gel-like substance of animal connective tissue; it supposedly serves as a lubricant and shock absorbent in the joints. Hyaluronic acid has also been isolated from umbilical cord, synovial fluid, skin, certain fowl tumors, and other sources. Treatment of this polysaccharide with the enzyme hyaluronidase, followed by acid hydrolysis, yields a disaccharide consisting of *N*-acetyl-D-glucosamine and D-glucuronic acid. This disaccharide appears to be the basic repeating structural unit that constitutes the hyaluronic acid molecule. The linkage between the two monosaccharide units in the disaccharide is of the



Repeating unit in the hyaluronic acid molecule

β type, involving carbon atom 1 (glycosidic hydroxyl) of the *N*-acetyl-D-glucosamine and carbon atom 4 of the D-glucuronic acid. Carbon atom 1 of the D-glucuronic acid in turn is attached through carbon atom 3 to another *N*-acetyl-D-glucosamine unit in the polysaccharide chain. Hyaluronic acid is thus a straight chain polymer composed of alternating β-1,3-glycosidic and β-1,4-glycosamidic linkages. The molecular weight of umbilical hyaluronic acid has been reported to be $3\text{–}8 \times 10^6$. *See* HYALURONIDASE; POLYSACCHARIDE. [W.Z.H.]

Hyaluronidase

An enzyme produced by mammals, reptiles, insects, and bacteria, which plays a part in such diverse mechanisms as fertilization of ova and bacterial invasion, and, because of its unique "spreading reaction," is used clinically to hasten fluid absorption. The spreading factor (SF) of F. Duran-Reynals, first discovered in testicle extracts, greatly enhanced spread of vaccine virus injected into skin. Such extracts also greatly accelerate the spread in skin of substances such as india ink, which, when mixed with SF, may diffuse within a few minutes giving a large stained area, whereas ink injected alone produces only a small bleb. The area of spread is determined by the concentration of SF injected and also by the volume and pressure of injection.

Pathogenic bacteria such as streptococci, pneumococci, and clostridia also secrete a SF. The name hyaluronidase was first applied to bacterial enzymes which depolymerized hyaluronic acid from bacterial capsules. The similarity of spreading factors from bacteria and testes, together with the isolation of hyaluronic acid from skin, identified testicular SF with hyaluronidase. The spreading reaction in skin is due to depolymerization by hyaluronidase of ground substance (hyaluronic acid gel) between cells. Dispersion of the follicle cells surrounding mammalian ova is based on a similar mechanism.

The substrate, hyaluronic acid, is a polymer of *N*-acetylglucosamine and glucuronic acid, occurring in umbilical cord, skin, synovial fluid, the vitreous humor of the eye, bacterial capsules, and certain tumors. Umbilical cord preparations in particular show high viscosity which is rapidly lost on incubation with hyaluronidase. Oligosaccharides or disaccharides, which give characteristic color reactions, are released as final products of stepwise hydrolysis by testis or bacterial enzymes respectively.

Assay of the hyaluronidase concentration can be by the spreading reaction, or more accurately by measuring the loss of viscosity on incubating enzyme and substrate under controlled conditions. In the turbidity reducing assay, dilutions of enzyme are incubated with substrate for a standard time and then acidified horse serum is added. The resulting turbidities are read and activity is expressed in turbidity reducing units.

Purification of hyaluronidase from testes or bacterial filtrates consists of fractionation with ammonium sulfate and alcohol.

Medical use includes injection to reduce swelling around bruises or fractures, and mixing with local anesthetics to accelerate diffusion. Infusion of fluid into tissues (clysis) is greatly facilitated by small amounts of hyaluronidase. Some success has been claimed in treating human infertility with the enzyme. *See* BACILLACEAE; BACTERIOLOGY, MEDICAL; PNEUMOCOCCUS; STAPHYLOCOCCUS; STREPTOCOCCUS; VIRULENCE. [J.F.M.]

Bibliography: F. Duran-Reynals (ed.), The ground substance of the mesenchyme and hyaluronidase, *Ann. N.Y. Acad. Sci.*, 52:943–1196, 1950.

The hydra, *Hydra littoralis*, body length to 25 mm. (From *E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

Hybodontoidea

An extinct group of sharks, prominent in late Paleozoic and early Mesozoic times. The Hybodontoidea are the most primitive of known sharks except for the Cladoselachii of the Devonian, from which they differ by the presence of claspers in the males and a narrow-based, more typical development of the paired fins. They are the only selachian shark

Primitive selachian, *Hybodus hauffianus*, a Mesozoic shark, original about 7½ ft long. (*After Smith Woodward*)

group represented in pre-Jurassic deposits; they appear to have arisen from the cladoselachians at the end of the Devonian and are abundantly represented, principally by teeth and spines, in the Carboniferous. They survived in diminished numbers in the Permian, flourished again in the Triassic, and then declined again, becoming extinct in the Eocene. *See* CLADOSELACHII; ELASMOBRANCHII. [A.S.R.]

Hydra

The type genus of the family Hydridae, class Hydrozoa, phylum Coelenterata. There are 11 species in the family in the United States, and 10 in the genus *Hydra*.

The hydras are almost universally studied in biology classes because of their simple structure and life history. They are among the simplest of diploblastic animals, representing the sac-within-a-sac stage of development similar to the gastrula of higher forms. Hydras are also frequently used in regeneration experiments.

Hydras vary in length from 2 to 25 mm. They are found attached to the bottom of fresh water bodies or on stems and sticks. They occur most commonly in quiet, shallow water, but may sometimes be found in swift water, or at considerable depth. Two species of hydra are world-wide in distribution; the others are limited to the North American continent or parts of it.

The hydra has an outer layer, the ectoderm, and an inner layer, the endoderm (sometimes called the gastrodermis). Between the two layers there is a gelatinlike substance, the mesoglea. The body is cylindrical and is attached at the posterior end by the basal disk. The mouth is on a raised area, the hypostome, and is surrounded by 4–12, usually 5 or 6, tentacles. The mouth leads into the gastrovascular cavity, comparable to the gastrocoel or archenteron of the gastrula state of higher animals. A primitive type of nervous system, the nerve net, connects each cell to its neighbors. Epitheliomuscular cells make up the bulk of the epidermis. They have a T-shaped base, terminating in contractile fibrils which contact the mesoglea. These fibrils form the longitudinal muscles. The epitheliodigestive cells on the inner layer are similarily equipped with fibrils which form the circular muscles. There are sensory cells in the outer layer, and interstitial cells in both. The nerve cells lie primarily in the mesoglea. Nematocysts, or sting cells, occur throughout the epidermis except on the basal disk. They are most numerous on the tentacles. Each nematocyst lies in a modified interstitial cell, the cnidoblast. Nematocysts are used both for defense and in securing food. Buds and sex cells also arise from interstitial cells. Gland cells are most numerous in the basal disk, where they produce an adhesive that attaches the hydra to the substrate. The gland cells may also produce a gas bubble by means of which the hydra can float in the water.

Hydras eat all sorts of small animals including larval fishes. Food is subdued by the nematocysts and brought to the mouth by the tentacles. Preliminary extracellular digestion occurs in the gastrovascular cavity, and the final stages of digestion occur in the adjacent cells. Food is passed through the cell membranes to those cells not in contact with the cavity. Undigested wastes are discharged through the mouth.

Locomotion is by somersaulting end over end, by inch-worming, sometimes called walking, and by floating.

Reproduction is commonly by budding. A new young hydra appears on the side of the parent, having at first a common gastrovascular cavity with it, but ultimately maturing and dropping off. Hydras also reproduce by longitudinal and transverse fission, and by sexual reproduction, which occurs primarily in the late fall or early winter. Hydras may be of separate sexes or hermaphroditic. If the latter, they have testes on their distal portion; ovaries are proximal. A cluster of interstitial cells merges to form a single ovum within the spherical swelling called the ovary. Quantities of sperm are produced in similar testes. Males far outnumber females when sexes are separate. Eggs are usually developed one at a time. Fertilization may be by the same or by another hydra. When the ovary is mature it ruptures, exposing the ovum which remains attached throughout fertilization and early development. The ovum drops off to complete development. There is no medusa stage or alternation of generations in the hydra. *See* HYDROZOA. [J.D.B.]

Hydrate

A particular form of a solid compound which has water in the form of H_2O molecules associated with it. For example, anhydrous copper sulfate is a white solid with the formula $CuSO_4$. When crystallized from water, a blue crystalline solid which contains water molecules as part of the crystals is formed.

Analysis shows that the water is present in a definite amount and the hydrate may be given the formula $CuSO_4 \cdot 5H_2O$. The water can be driven off by gentle heating to give the original anhydrous white solid.

The water is held by different means in these compounds. In $CuSO_4 \cdot 5H_2O$, four of the water molecules are attached to the copper ion in the manner of coordination complexes. The fifth water molecule is related to the sulfate and presumably held ...

... with the anion or cation. The water occupies a definite place in the crystal lattice. Alums, with their 12 molecules of water, are examples of this.

Some crystalline compounds contain water which is not present in definite proportions. Hydrates of this kind are *zeolites* and other silicate minerals. Some metallic insoluble hydroxides may actually be hydrates. *See* COORDINATION CHEMISTRY; EQUILIBRIUM, PHASE; HYDRATION; HYDROXIDE. [E.E.WR.]

Hydration

The incorporation ...

... by relatively weak forces or may exist as a definite compound.

Many salts form solid hydrates when exposed to water vapor under certain conditions of temperature and pressure. Copper sulfate, for example, forms a monohydrate ($CuSO_4 \cdot H_2O$) when exposed at 25°C to water vapor at a pressure of 0.8 mm of mercury. At higher pressures, other hydrates are formed. Water is lost from these compounds when they are heated or when the water vapor pressure falls below a minimum value. Solids forming hydrates at low pressures are used as drying agents. *See* DELIQUESCENCE; DESICCANT; EFFLORESCENCE; SOLUTION; SOLVATION. [F.J.J.]

Hydraulic actuator

A form of hydraulic motor that produces linear motion. An actuator receives oil or other hydraulic fluid under pressure from a pump. The oil pressure times the piston area of the actuator gives the force developed (*see* HYDRAULIC PRESS). Cylinder pressures are usually 500–3000 psi. Piston area is commonly 1¾–300 in.² For the push stroke, the oil operates against the full piston area; for the pull stroke of a double-acting unit, the oil operates against that portion of the piston beyond the piston rod. Rod area is ⅕–24 in.² In a single-acting unit, a spring returns the rod. A single-acting unit is usually limited by the spring to a stroke less than a foot; a double-acting unit may have as long a stroke as the stiffness of the rod permits, commercially up to 24 ft.

Basic hydraulic control. An hydraulic actuator may be used to move any mechanism. Typical applications are to machine tools, construction machinery, aerodynamic control surfaces, and steering mechanisms (*see* STEERING, POWER).

As shown in Fig. 1, oil from a storage tank is pumped to high pressure and directed to the appropriate side of a double-acting hydraulic actuator. The cylinder of the actuator may be rigidly fastened to one member such as the bed of a ma-

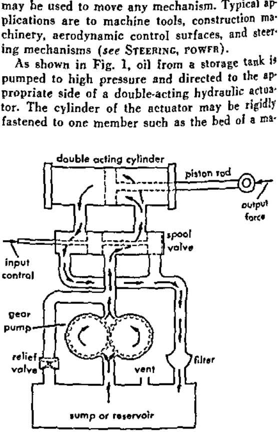


Fig. 1. Elements of a hydraulic control.

chine, and the piston rod attached to another member, such as a tool feed, that is to move relative to the first. For long trouble-free life of the actuator, it is essential that only axial forces react on the piston. Thus the cylinder is mounted on a pivot if the piston end follows a curved path.

A spool valve or other control means admits oil to the desired side of the actuator cylinder and opens a return path for oil from the low-pressure side. For rapid operation, the valve and piping should present little resistance to the flow of oil. A filter in the return line may be desirable to reduce wear of the closely fitting parts of pump, valve, and actuator by trapping scratch particles from the moving parts.

The pump is the source of power for the system. For simple motive applications where the actuator serves to apply a large force in accordance with a small, usually manual, control force, the pump has fixed displacement. A spring-loaded relief valve bypasses the hydraulic fluid after operating pressure is reached. For follow-up systems as in automatic controls, a variable-displacement pump may be used; in such pumps, a variable eccentric adjusts the length of piston stroke. Pump adjustment is then used in addition to or in place of valve control.

Actuator characteristics. The hydraulic actuator is a rigid cylinder containing a closely fitting piston, as shown in Fig. 2. The piston rod projects from a cylinder head through a tightly fitting bushing. A cap closes the opposite end of the cylinder; tie rods hold cap and head against the internal hydraulic pressure

The parts set in motion by the actuator may have appreciable mass. For this reason, the actuator includes a damping feature to bring the piston and its connected load smoothly to rest. This feature consists of a recess into which the piston runs at the end of its travel. The entering portion of the piston may taper to cut off gradually the escape of oil from the recess, several escape passages may be arranged to close successively as the piston advances into the recess, or as in Fig. 2, an escape passage may contain a spring-loaded check ball. A second passage with an adjustable leak may provide additional control of the cushioning.

Where space and weight are critical, actuators are operated at high pressure. Where space and weight are not critical, low-pressure units are usually more economical. Should space be limited in some portions of a system and not in others, a low-

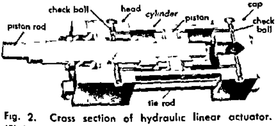


Fig. 2. Cross section of hydraulic linear actuator. (*Flick-Ready Corp.*)

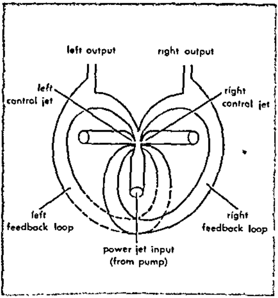


Fig. 3. In an all-fluid valve, control jets deflect the power stream; a portion of the power stream can be fed back to hold the power jet in the last output called for.

pressure system can be used with boosters. Boosters are actuators with large pistons that drive single-stroke pumps (actuators operated in reverse) with smaller pistons. In this way, low-pressure oil in the main system provides high-pressure oil in a branch line to operate a small high-pressure actuator located in close quarters. *See* DISPLACEMENT PUMP; SERVOMECHANISM.

Fluid amplifier. An equivalent experimental valve with no moving parts is shown in Fig. 3. Either oil or air can serve as the fluid. When used in the arrangement described above, a pump delivers fluid under pressure for the output, which is an actuator for translation or possibly a turbine for rotation. The fluid stream is deflected in this all-fluid valve by control jets having as little as $^1/_{100}$ the energy of the power stream. Response frequency of this fluid amplifier in small sizes can be as high as 20,000 cps; linearity is comparable to that of a spool valve; it can be designed for a wide range of power outputs. [F.H.R.]

Bibliography: W. Ernst, *Oil Hydraulic Power and Its Industrial Applications*, 2d ed., 1960; J. C. Truxal (ed.), *Control Engineers' Handbook*, 1958.

Hydraulic analog table

An experimental facility which makes use of the analogy between water flow with a free surface and two-dimensional compressible gas flow. The water flows over a smooth horizontal surface and is bounded by vertical walls geometrically similar to the boundaries of the corresponding compressible gas flow. The change in water depth between a reference station and any other point in the flow is related to the change in density, pressure, and tem-

perature in the analogous gas flow. The analog table is an effective low-cost research tool; flow patterns are easily observed and boundary changes may be made rapidly and comparatively inexpensively during exploratory studies.

Theoretical basis. The basic assumptions are that both liquid and gas flows are inviscid (no energy dissipation) and that the vertical pressure distribution in the free surface flow is hydrostatic. Under these conditions the analogy may be obtained by writing the energy and continuity equations for the water and gas.

From the energy equation for water, the velocity is $V = \sqrt{2g(y_0 - y)}$ and $V_{max} = \sqrt{2gy_0}$ where y is the water depth and the subscript $_0$ refers to the value at a stagnation point ($V = 0$).

The corresponding equations for a gas are

$$V = \sqrt{2gc_p(T_0 - T)} \quad \text{and} \quad V_{max} = \sqrt{2gc_pT_0}$$

where c_p is the specific heat at constant pressure and T is the temperature.

If the ratios V/V_{max} are equated for water and gas

$$\frac{y_0 - y}{y_0} = \frac{T_0 - T}{T_0} \quad \text{or} \quad \frac{y}{y_0} = \frac{T}{T_0}$$

If u and v are the x and y components of velocity V, the continuity equations for water and gas are given by

$$\frac{\partial(uy)}{\partial x} + \frac{\partial(vy)}{\partial y} = 0 \quad \text{(water)}$$

$$\frac{\partial(u\rho)}{\partial x} + \frac{\partial(v\rho)}{\partial y} = 0 \quad \text{(gas)}$$

Hence, it follows that depth y and density ρ are also analogous quantities, thus

$$\frac{y}{y_0} = \frac{\rho}{\rho_0} = \frac{T}{T_0}$$

However, in an isentropic gas flow

$$\frac{\rho}{\rho_0} = \left[\frac{T}{T_0}\right]^{1/(k-1)} = \left[\frac{p}{p_0}\right]^{1/k}$$

Therefore, the analogy requires that the adiabatic exponent $k = 2$.

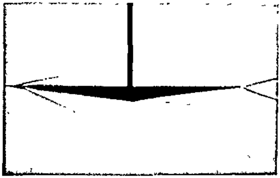
The relations of the classical analogy may therefore be summarized as

$$\frac{y}{y_0} = \frac{\rho}{\rho_0} = \frac{T}{T_0} = \sqrt{\frac{p}{p_0}}$$

In water the speed of propagation of a small disturbance is $c = \sqrt{gy}$, and in a gas the corresponding quantity is the acoustic velocity $c = \sqrt{kp/\rho}$.

The ratio $V/\sqrt{gy}$ is known as the Froude number in water and in a gas the analogous ratio $V/\sqrt{kp/\rho}$ is the Mach number.

Applications. An analogy has been shown to exist for a gas with $k = 2$; however the applications are primarily for air where $k = 1.4$. Various proposed



Shadowgraph of diamond-shaped airfoil in hydraulic analog table.

modifications to the classical analogy have shown the analog table to be a quantitatively useful tool for many aerodynamic problems.

In the realm of transonic flow, experimental results from the water table can be interpreted for a hypothetical gas with $k = 2$ and transferred to air by means of the transonic similarity laws.

In supersonic flow, the classical analogy is no longer strictly valid because the presence of shock waves invalidates the condition of negligible energy dissipation. For the indicated classical relationships between the water depth and the gas density, temperature, and pressure, only the depth and density relation remains useful for supersonic air flow.

If the air density distribution is obtained from water-table experiments, the corresponding pressure or temperature distributions can be calculated with good accuracy using the appropriate aerodynamic equations. Pressure distributions around supersonic airfoils and cascades have been studied [illegible]

objects. These water-table diffraction investigations are the counterpart of shock-tube studies in air. See WIND TUNNEL. [D.R.F.H.]

Bibliography: A. T. Ippen and D. R. F. Harleman, *Certain Quantitative Results of the Hydraulic Analogy to Supersonic Flow*, Ohio State Univ. Eng Exp. Sta. Bull. 149, 1952.

Hydraulic gradient

The slope along a closed water conduit that measures the sum of elevation and pressure heads. Total head consists of these heads plus the velocity head (*see* BERNOULLI'S THEOREM). The elevation and pressure heads at a point can be measured directly by connecting a tube to a small hole in the conduit wall and observing the vertical height to which the water rises in the tube, called a piezometer. [D.R.F.H.]

Hydraulic jump

An abrupt increase of depth in a free-surface liquid flow. A hydraulic jump is characterized by rapid flow and small depths on the upstream side

and by larger depths and smaller velocities on the downstream side. A jump can form only when the upstream flow is supercritical; that is, when the fluid velocity is greater than the propagation velocity c of a small, shallow-water gravity wave ($c = \sqrt{gh}$, where h is the depth). A considerable amount of energy is dissipated in the conversion from supercritical to subcritical flow. *See* CHANNEL, OPEN; WAVE MOTION IN LIQUIDS. [D.R.F.H.]

Hydraulic press

A combination of a large and a small cylinder connected by a pipe and filled with a fluid so that the pressure created in the fluid by a small force acting on the piston in the small cylinder will result in a large force on the large piston. The operation depends upon Pascal's principle which states that when a liquid is at rest the addition of a pressure (force per unit area) at one point results in an identical increase in pressure at all points. Therefore, in Fig. 1, the pressure due to the application of force F_1 is

$$p = F_1/A_1$$

and the equilibrium force F_2 is

$$F_2 = pA_2 = F_1\frac{A_2}{A_1}$$

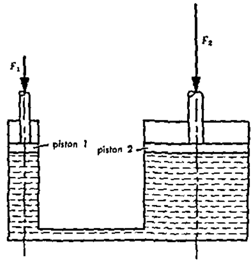


Fig. 1. *Principle of hydraulic press.*

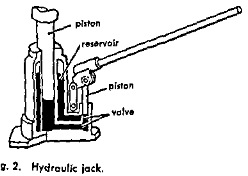


Fig. 2. Hydraulic jack.

where A_1 and A_2 are the areas of pistons 1 and 2 respectively. The mechanical advantage is

$$\text{MA} = \frac{F_2}{F_1} = \frac{A_2}{A_1}$$

The principle of the hydraulic press is used in lift jacks, earth-moving machines, and metal-forming presses (Fig. 2). A comparatively small supply pump creates pressure in the hydraulic fluid. The fluid then acts on a substantially larger piston to produce the action force. In this way forces greater than 15,000 tons are developed over the entire stroke of hydraulic presses. Heavy objects are accurately weighed on hydraulic scales in which precision ground pistons introduce negligible friction. *See* MECHANICAL ADVANTAGE; SIMPLE MACHINE.

[R.M.PH.]

Hydraulic turbine

A machine which converts the energy of an elevated water supply into mechanical energy of a rotating shaft. Most old-style water wheels utilized the weight effect of the water directly, but all modern hydraulic turbines are a form of fluid dynamic machinery of the jet and vane type operating on the impulse or reaction principle and thus involving the conversion of pressure energy to kinetic energy. The shaft invariably drives an electric generator, and speed accordingly must be of an acceptable synchronous value.

The impulse or Pelton unit has all available energy converted to the kinetic form in a few stationary nozzles and subsequent absorption by reversing buckets mounted on the rim of a wheel (Fig. 1). Reaction units of the Francis or the Kaplan types run full of water, submerged, with a draft tube and a continuous column of water from head race to tail race (Figs. 2 and 3). There is some fluid acceleration in a continuous ring of stationary nozzles with full peripheral admission to the moving nozzles of the runner and in which there is further acceleration. The draft tube produces a negative pressure in the runner with the propeller or Kaplan units acting as suction runners; the Francis inward-flow units act as pressure runners. *Mixed-flow* units give intermediate degrees of rotor pressure drop and fluid acceleration. Reaction units use vertical shafts for better accommodation of the draft tube, whereas Pelton units usually have a horizontal shaft. Kaplan units employ adjustable propeller blades as well as adjustable stationary nozzles in the gate ring for higher sustained efficiency (Fig. 4). Pelton units are preferred for high head service (about 1000± ft), Francis runners for medium heads (500± ft), and propeller or Kaplan units for low heads (100± ft).

Hydraulic-turbine performance is rigorously defined by characteristic curves, such as the efficiency [illegible]

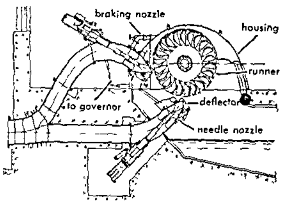


Fig. 1 Cross section of an impulse (Pelton) type of hydraulic-turbine installation. (From *T. Baumeister, ed., Marks' Mechanical Engineers' Handbook, 6th ed., McGraw-Hill, 1958*)

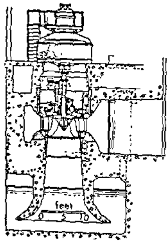


Fig 2. Cross section of a reaction (Francis) type of hydraulic turbine installation. (From *J. H. Perry, ed., Chemical Engineers' Handbook, 3d ed., McGraw-Hill, 1950*)

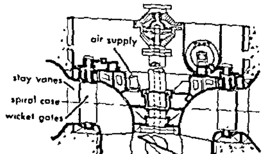


Fig. 3. *Cross section of a propeller (Kaplan) type of hydraulic turbine. (T. Baumeister, ed., Marks' Mechanical Engineers' Handbook, 6th ed., McGraw-Hill)*

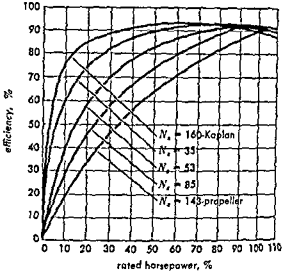


Fig 4. Efficiency characteristics of selected hydraulic turbine types. (From *J. H. Perry, ed., Chemical Engineers' Handbook, 3d ed., McGraw-Hill, 1950*)

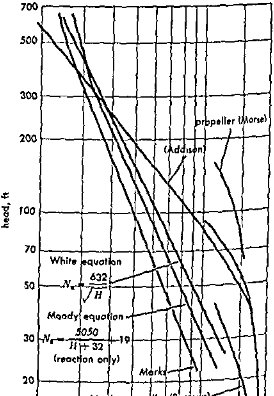


Fig. 5. Hydraulic turbine experience curves; specific speed vs. head. (From *J. H. Perry, ed., Chemical Engineers' Handbook, 3d ed., McGraw-Hill, 1958*)

site head as the result of accumulated experience on satisfactory turbine installations. Specific speed N_s is a criterion or coefficient which is uniquely applicable to a given turbine type and relates head, power, and speed, which are the basic performance data in the selection of any hydraulic turbine. Specific speed is defined as

$$N_s = \frac{\text{rpm} \times (\text{shp})^{0.5}}{(\text{head})^{1.25}}$$

where rpm is revolutions per minute, shp is shaft horsepower, and head is head on unit in ft. Specific speed is usually identified for a particular unit at the point of maximum efficiency. Cavitation must also be carefully scrutinized in any practical selection.

The draft tube (Fig. 2) is a closed conduit which (1) permits the runner to be set safely above tail water level yet to utilize the full head on the site, and (2) is limited by the atmospheric water column to a height substantially less than 30 ft, and when made flaring in cross section will serve to recover velocity head and to utilize the full site head.

Efficiency of hydraulic turbine installations is always high: more than 85% after all allowances for hydraulic, shock, bearing, friction, generator, and mechanical losses. Material selection is not only a problem of machine design and stress loading from running speeds and hydraulic surges but is also a matter of fabrication, maintenance, and resistance to erosion, corrosion, and cavitation pitting.

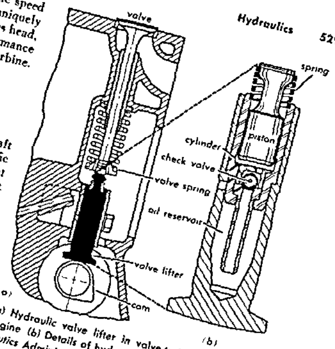
Governing problems are severe, primarily because of the large masses of water involved and their positive and negative acceleration without interruption of the fluid column continuity and the consequent shock and water-hammer hazards. See PRIME MOVER. [T.B.]

Bibliography: H. K. Barrows, *Water Power Engineering*, 3d ed., 1943; T. Baumeister, *Marks' Mechanical Engineers' Handbook*, 6th ed., 1958; W. P. Creager and J. B. Justin, *Hydroelectric Handbook*, 2d ed., 1950.

Hydraulic valve lifter

A device that eliminates the need for mechanical clearance in the valve train of internal combustion engines.

Clearance is normally required to prevent the valve being held open and destroyed as the valve train undergoes thermal expansion. However, clearance requires frequent adjustment and is responsible for much operating noise. The hydraulic lifter is a telescoping compression strut in the linkage between cam and valve, consisting of a piston and cylinder. When no opening load exists, a weak spring moves the piston, extending the strut and eliminating any clearance. This action sucks oil into the cylinder past a check valve. The trapped oil transmits the valve opening forces with little deflection. A slight leakage of oil during lift shortens the strut, assuring valve closure. The leakage oil is replaced as the spring again extends the strut at no load. See VALVE TRAIN. [A.R.B.]



(a) Hydraulic valve lifter in valve-train of an L-head engine (b) Details of hydraulic valve lifter. (Civil Aeronautics Administration and Texas Co.)

Hydraulics

The behavior of water or other liquids, chiefly when in motion. As a part of fluid mechanics, hydraulics deals with the properties, behavior, and effects of all liquids at rest against or in motion relative to boundary surfaces or objects. Its laws apply also to gases when their density changes, or when thermal effects or velocities are small and their compressibilities negligible. See FLUID MECHANICS.

Applications of hydraulics. Common hydraulic applications encountered in civil engineering include the flow of water in pipes, canals, and rivers, in flood control, in land reclamation, and in hydroelectric power projects. Hydraulics also influences appurtenant structures and devices such as dams, dikes, locks, spillways, weirs, piers, ships' hulls, gates, and valves. Mechanical and chemical engineering applications include the flow of gases, oils, or other liquids, lubrication, fluid machinery such as pumps, turbines, propellers, fans, and fluid power transmission and control devices, including servomechanisms.

The physical laws of hydrostatics govern the effects of fluids at rest. Hydrokinematics covers fluid motion wherein consideration of the forces causing the motion is not required. Hydrodynamics treats of fluid motion, the forces involved, and the accompanying energy changes. Characteristics of particular phenomena in hydraulics include fluid properties; the shapes, sizes, relative roughnesses, and

relative motions of the surfaces or objects involved; the relative times and distances (short or long); whether flow is closed as in a pipe or open as in a canal; whether motion is in a relatively extensive fluid as a moving submarine deeply submerged, or whether motion occurs where there is a free surface as in the case of a ship at sea.

Physical laws are expressed by mathematical equations; two are required—an equation of condition (continuity) and an equation of motion (momentum or energy) When introduced by methods of dimensional analysis, physical quantities which are known, assumed, or sought experimentally may enter the equations in combinations which are dimensionless numbers or ratios.

Properties. Hydraulics deals with bulk fluids which are regarded as homogeneous; it is not concerned with molecular sizes. As distinct from solids, fluids are unable to resist shearing forces while remaining in a state of equilibrium. Among fluids, a liquid mass will have a definite volume, varying only slightly with temperature, and pressure, whereas a mass of gas will fill any space *available to it*, the pressure adjusting itself accordingly.

A common fluid property is density, the mass per unit volume, usually expressed dimensionally in slugs, or pound-seconds squared divided by feet to the fourth power. Specific weight, the weight per unit volume in pounds per cubic foot, is another property

Viscosity, the ability of a fluid in motion to resist shearing forces, is the seat of all fluid resistance. The coefficient of viscosity (absolute viscosity) is the intensity of the shearing stress divided by the rate at which adjacent fluid layers slip past each other; it has the dimension slugs per foot-second or pound-seconds per square foot. The slippage rate is greatest adjacent to a solid boundary. There, in the boundary layer, the velocity rises sharply from zero at the wall toward the value in the main stream. Where viscosity and density enter a problem, the coefficient of viscosity is divided by density to give the kinematic viscosity, which has the dimension square feet per second.

Elasticity denotes the relative compressibility of a fluid under pressure. The volume modulus of elasticity is the change in pressure intensity divided by the corresponding unit volumetric change. This is not important in ordinary applications but plays a significant role when gases flow at high velocity or when, in a pipeline carrying a liquid, a valve is closed or opened rapidly, thereby causing a pressure wave. *See* WATER HAMMER.

Surface tension is a force per unit of length that binds surface molecules of a liquid to one another or to a solid boundary for which they may have an affinity; it is of minor importance except where thin films with a free surface exist or capillary movements such as that of water through soils occur. *See* DENSITY; ELASTICITY; HYDROSTATICS; SURFACE TENSION; VISCOSITY OF LIQUIDS.

[W.A.]

Hydrazine

A colorless liquid, H_2NNH_2 (boiling point 114°C), with a musty, ammonialike odor. Physically it is similar to water, but chemically it is reducing, decomposable, basic, and bifunctional. Its derivatives range from simple salts to ring compounds, polymers, and coordination complexes. Major uses of hydrazine include such diverse applications as rocket fuels (since combustion of hydrazine is highly exothermic), corrosion inhibition in boilers, and syntheses of biologically active materials.

Hydrazine is manufactured by two routes: the reaction of chloramine with ammonia, and the reaction of sodium hypochlorite with urea. Both processes require the presence of glue or gelatin to inhibit catalytic decomposition of the product by unreacted oxidants. Because hydrazine forms an azeotrope containing 31% water (boiling point 121°C), anhydrous hydrazine is isolated from aqueous process streams by extractive distillation with aniline

Added to feed water, hydrazine reduces rust in boilers to a hard film of magnetic iron oxide and reduces oxygen to water on catalytic metal surfaces. Metal ions such as Cu^{++} and Ni^{++} are reduced to free metals, and organic nitro compounds are reduced to amines by hydrazine. The energetic reaction of hydrazine with strong oxidants such as nitric acid is utilized in rocket propulsion. The thermal decomposition of hydrazine produces free radicals and gases, useful in rubber *curing* and foam-rubber production.

A slightly weaker base than ammonia, hydrazine forms most of the analogs of ammonia derivatives as well as distinctive hydrazine derivatives in which both nitrogen atoms are involved. For example, hydrazine forms two series of salts, such as $N_2H_4{\cdot}HCl$ and $N_2H_4{\cdot}2HCl$, and forms not only *hydrazones*, $RCH{=}NNH_2$, but also azines, $RCH{=}NN{=}CHR$, by reaction with aldehydes. In a manner similar to ammonia, hydrazine attacks polar bonds, in one case displacing ammonia from urea to form semicarbazide, a reaction which can be reversed by an excess of ammonia.

Prominent uses of hydrazine derivatives include rocket fuels [1,1-dimethylhydrazine, $(CH_3)_2NNH_2$], *antituberculin drugs* (isonicotinic hydrazide, $C_5H_4N{\cdot}CO{\cdot}NHNH_2$), plant-growth regulators (maleic hydrazide, $NH{\cdot}CO{\cdot}CH{=}CH{\cdot}CO{\cdot}NH$, and β-hydroxyethyl hydrazine, $HOCH_2CH_2{\cdot}NHNH_2$), dye and explosive intermediates [aminoguanidine, $NH_2{\cdot}C(NH){\cdot}NHNH_2$], algaecides and fungicides [copper dihydrazinium sulfate, $CuSO_4{\cdot}(N_2H_5)_2SO_4$], soldering fluxes (hydrazine hydrobromide, $N_2H_4{\cdot}HBr$), blowing agents for foam rubber (azides, $R{\cdot}CO{\cdot}N_3$, and sulfonyl hydrazides, $R{\cdot}SO_2{\cdot}NHNH_2$), insecticides (1,4-diphenylsemicarbazide, $C_6H_5{\cdot}NH{\cdot}CO{\cdot}NHNH{\cdot}C_6H_5$), heterocycle syntheses (thiosemicarbazide, $NH_2{\cdot}CS{\cdot}NHNH_2$), and polymers (dihydrazide-formaldehyde resins). *See* AMMONIA; NITROGEN. [T.H.D.]

Bibliography: L. F. Audrieth and B. A. Ogg, *The Chemistry of Hydrazine*, 1951.

Hydride

A compound containing hydrogen and another element. Many of the elements combine with hydrogen to form compounds, and even though hydrogen is less negative than the second element in the compound, they are called hydrides. For example, H_2S is called a hydride in the discussion here although it is usually called hydrogen sulfide and properly so. Hydrides can be divided into three classes of compounds: covalent hydrides, saltlike hydrides, and metallic hydrides.

Covalent hydrides refer to compounds such as H_2O, H_2S, and NH_3, which are volatile. Covalent hydrides are formed from the nonmetals. Carbon hydrides would refer to natural gas, CH_4, and other hydrocarbons containing chains of carbons. Boron hydrides also may contain more than one boron atom and have some importance as rocket fuels. Saltlike hydrides are ionic in nature and nonvolatile. Lithium hydride, LiH, is a typical example. These compounds contain negative hydrogen ions which migrate to the positive pole when molten compounds are electrolyzed. These compounds are used as sources of hydrogen, as drying agents, and in organic syntheses.

Metallic hydrides include compounds with simple formulas such as uranium hydride, UH_3, or palladium hydride, PdH_x. Palladium metal will absorb various amounts of hydrogen, depending upon the temperature. Some of these compounds are thought to contain the hydrogen in the holes in the metallic lattice. They retain many metallic characteristics. See HYDRIDE, METAL; HYDROCARBON; HYDROGEN. [E.E.WR.]

Hydride, metal

A compound in which hydrogen is bonded chemically to a metal or metalloid element. Metal hydrides may be classified generally as ionic hydrides, covalent hydrides, and transitional metal hydrides. The periodic table of the elements can be arranged according to the type of hydride formed. Complex hydrides, containing more than one metallic component, also are known.

Metal hydrides are used as reducing agents, as agents for producing or purifying hydrogen, as high-energy fuels, as intermediates in the synthesis of other compounds, and as reagents for the deposition of metals. Complex hydrides exhibit unique and selective reducing powers, making them valuable in certain syntheses such as the preparation of vitamins. However, many of the metal hydrides are of academic interest only.

Ionic hydrides. In the ionic hydrides of the alkali and alkaline earth metals in groups Ia and IIa, hydrogen is present as the negatively charged hydride ion, H^-, which is formally analogous to a halide ion such as fluoride or chloride. Thus, the ionic hydrides are saltlike solids with simple formulas (NaH, CaH_2), high melting points, high heats of formation, and a high degree of thermal stability. They are insoluble in organic solvents and are able to conduct electricity if they are melted. If a molten ionic hydride such as lithium hydride is electrolyzed, hydrogen is evolved at the positive electrode. The usual method of preparing an ionic hydride is to treat the metal with hydrogen under pressure at an elevated temperature.

The ionic hydrides are strong reducing agents and are highly reactive. They are used as sources of hydrogen. They react exothermically, some of them violently, with water to liberate hydrogen gas. Alkali metal hydrides may ignite spontaneously in moist air. Calcium hydride is used as a portable source of hydrogen and as a reducing agent in the preparation of certain metals such as zirconium. Sodium hydride finds use as a catalyst for organic condensation reactions, as a descaling agent in the pickling of steel, and as an intermedi- [illegible]

energy fuels and complex hydride reducing agents such as lithium aluminum hydride.

Covalent hydrides. The covalent hydrides of the metals in groups IIIb through VIb are volatile gases, liquids, and solids in which hydrogen is bonded to the parent element by shared electron-pair bonds. Generally they are soluble in organic solvents, have relatively low heats of formation, low melting and boiling points, and low thermal stabilities. The chemical behavior of hydrogen in these compounds is much like that of the methyl group in corresponding organometallic compounds. The chemical formulas of the simple covalent hy-

Table 1. Long-form periodic table of the elements showing distribution of hydride types

Group	Ia	IIa	IIIa	IVa	Va	VIa	VIIa	VIIIa			Ib	IIb	IIIb	IVb	Vb	VIb	VIIb	0
	H																	He
	Li	Be											B	C	N	O	F	Ne
	Na	Mg											Al	Si	P	S	Cl	Ar
	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
	Cs	Ba	57–71	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
	Fr	Ra	89–96															
	Ionic hydrides		Transitional metal hydrides								Intermediate hydrides			Covalent hydrides				

Elements 57–71 (Rare-earth metals): La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu
Elements 89–96 (Actinide elements): Ac, Th, Pa, U, Np, Pu, Am, Cm, Bk, Cf, Es, Fm, Md, No

drides correspond to the primary valences of the parent elements (BH_3, SnH_4, SbH_3, TeH_2), but many of the elements in groups IIIb, IVb, Vb, and VIb form more than one hydride. Silicon and germanium form series analogous to the aliphatic hydrocarbons, SiH_4, Si_2H_6, Si_3H_8. However, decreasing stability with increasing molecular complexity limits these series to relatively few members. Dimeric or polymeric hydrides also are known for some of the elements in groups Vb and VIb, such as arsenic and selenium. The boron hydrides form an extensive system of compounds ranging from simple molecules to highly polymeric, nonvolatile solids, in which bonding can occur both by direct linking of boron atoms and by a type of chemical bond, unique to electron-deficient systems, known as hydrogen-bridge bonding. In diborane (B_2H_6), for example, the two boron atoms are bonded by a pair of hydrogen atoms shared between them, the other four hydrogens being bonded normally to boron. The hydrides of aluminum and gallium also are polymerized by this type of bonding. *See* HYDROGEN BOND.

The covalent hydrides are not nearly so reactive as the ionic hydrides. Most of them are rather weak reducing agents. They are readily oxidized; but only a few, notably the hydrides of silicon and certain of the boron hydrides, are spontaneously flammable. Only the group IIIb hydrides are decomposed readily by water. Most of the covalent hydrides will decompose at temperatures below a red heat, and with the heavier elements such as lead, bismuth, and tellurium, the thermal stability of the hydrides is relatively low.

As a general rule, covalent hydrides cannot be prepared by the direct reaction of metal with hydrogen. They are synthesized by the hydrolysis of an alloy, such as magnesium silicide, aluminum telluride, or sodium arsenide; or by the metathetical reaction of a metal halide such as boron trifluoride with an alkali metal hydride such as lithium hydride (or its [illegible]

The impor[illegible] in the preparation of special hydride reducing agents and for the production of high-energy fuels. Partially substituted derivatives of silicon hydride, such as dichlorosilane, SiH_2Cl_2, are used to some extent in the preparation of organosilicon polymers, or silicones.

The properties of the hydrides of beryllium, magnesium, and zinc appear to be intermediate between those of the ionic hydrides and the covalent hydrides.

Hydrides of the transitional metals. The hydrides of the transitional metals in groups IIIa through VIIIa exhibit wide differences in properties. The capacity of the various metals in this category to adsorb, absorb, occlude, or otherwise react with hydrogen decreases progressively with increasing group number. The elements in groups IIIa and IVa rank highest in their ability to absorb hydrogen, and the reaction of hydrogen with these metals is quite exothermic. At the other end of the scale, most of the metals in group VIIIa are practically inert to hydrogen. Palladium, however, is a notable exception and can dissolve large volumes of hydrogen.

The diffusion of hydrogen into many transitional metals is believed to occur primarily through minute defects in the metal structure. This process may be accompanied by compound formation or by solution of the hydrogen as atomic hydrogen in the metal. This process is believed to be analogous to alloy formation with some metals, such as palladium. The absorption of hydrogen causes an expansion of the metal crystal structure and, with the more reactive metals such as cerium and zirconium, a loss of metallic character. With less reactive metals such as iron, reaction may be confined to simple internal adsorption of hydrogen at grain boundaries and defects, and little if any metallic character is lost. The transitional metal hydrides, as usually prepared by heating the metals in hydrogen, almost never contain amounts of hydrogen corresponding to exact chemical formulas, and the ease with which hydrogen is lost is a function of the hydrogen content. With some metals, in-

Table 2. Physical properties of representative hydrides

Compound	Formula	Form	Melting point, °C	Boiling point, °C	Density, g/ml
Lithium hydride	LiH	White crystal	680°	Dissociation pressure, dp = 27 mm Hg at 680°	0.76
Sodium hydride	NaH	White crystal	Decomposes	dp = 1 atm at 425°	1.396
Calcium hydride	CaH_2	White crystal	>1000°	dp = 1 atm at 1000°	1.9
Diborane	B_2H_6	Colorless gas	−165.5°	−92.5°	0.438 at bp
Silane	SiH_4	Colorless gas	−185°	−111.8°	0.68 at mp
Stannane	SnH_4	Colorless gas	−150°	−52°	
Arsine	AsH_3	Colorless gas	−113.5°	−55°	
Stibine	SbH_3	Colorless gas	−88.5°	−17°	2.2 at bp
Tellurium hydride	TeH_2	Colorless gas	−51°	−4°	2.7 at −18°
Zirconium hydride	$ZrH_{1.n}$	Gray powder		Decomposes about 500°	5.47
Uranium hydride	UH_3	Black powder	Decomposes	dp = 32.5 mm at 307°	11.4

cluding iron and chromium as well as palladium, hydrogen may be occluded or dissolved during electrolytic reactions such as plating or electrolysis. The penetration and embrittlement of iron by hydrogen produced by corrosion or other electrolytic processes is a problem of considerable economic importance.

With the exception of the hydrides of the group IIIa elements, the transitional metal hydrides are relatively inert. Even the hydrides of titanium and zirconium appear no more reactive than the free metals in an equal state of subdivision.

Group IIIa includes the rare-earth elements (57–71) and the actinide elements (89–103). These metals absorb hydrogen avidly at moderate temperatures (200–500°C) and lose their metallic appearance, forming dark grey or black, powdery hydrides. Metal powders formed by the decomposition of previously formed hydrides will absorb hydrogen at room temperature. As usually prepared, the group IIIa hydrides do not have exact formulas and their dissociation pressures of hydrogen depend on hydrogen content. The rare earth hydrides are more stable toward thermal dissociation than the actinide hydrides. In a hydrogen atmosphere, cerium and lanthanum hydrides are not entirely dissociated at 1000°C, whereas uranium hydride, UH_3, has a dissociation pressure of more than one atmosphere at 440°C. The dissociation pressure of uranium deuteride is about forty percent greater than that of uranium hydride.

Although all of the metals in group IIIa will, presumably, form hydrides, relatively few have been prepared and studied. Principal compounds include the hydrides of cerium, lanthanum, thorium, uranium, and plutonium. The compound studied in greatest detail, uranium hydride, contains hydrogen-bridge bonds linking uranium atoms. This type of bonding may occur in others of the group IIIa hydrides. Uranium hydride has considerable metallic character and will conduct electricity.

The group IIIa hydrides are very reactive, and many of these compounds, including the hydrides of cerium, thorium, and uranium, will ignite spontaneously and oxidize upon exposure to air. They are very powerful reducing agents capable of reacting with a wide variety of reagents including water, acids, and the halogens, to form oxy or halo derivatives of the parent elements.

Because of their extreme reactivity, the rare-earth metal and actinide metal hydrides are not items of commerce. The use of uranium hydride, or deuteride, as a medium for the storage and delivery of very pure hydrogen is the basis of one laboratory method.

The hydrides of titanium and zirconium are used to a limited extent in metallurgy and for depositing films of these metals as part of a process for joining metals to ceramics. The ability of palladium to dissolve hydrogen is used in the purification of this gas.

From what little is known, the behavior of the group Ib metals toward hydrogen appears to be much like that of the transitional metals, although copper heated in hydrogen does form a gaseous, unstable hydride in the temperature region 400–500°C.

Under extreme conditions such as in electric discharges, many metals will form volatile, short-lived transient hydrides of the general formula type MH. Although some of these can be demonstrated experimentally, most are observed only by their appearance in spectra. They are important in studying the physics of molecular structure. The action of atomic hydrogen at low temperatures forms surface films of unstable hydrides with many metals. See HYDROGEN. [D.T.H.]

Bibliography: D. T. Hurd, *An Introduction to the Chemistry of the Hydrides*, 1952; E. G. Rochow, *Introduction to the Chemistry of the Silicones*, 2d ed., 1951; D. P. Smith, *Hydrogen in Metals*, 1948; A. E. Stock, *The Hydrides of Boron and Silicon*, 1933.

Hydrocarbon

One of a group of chemical compounds composed only of hydrogen and carbon. The very large variety of hydrocarbons can best be divided into three classes (aliphatic, alicyclic, and aromatic), each of which may be further divided into a number of subclasses.

Aliphatic hydrocarbons. These are open-chain compounds which may be saturated or unsaturated. The saturated compounds, known as paraffin hydrocarbons or alkanes, include methane and its homologs having the empirical formula C_nH_{2n+2}. The unsaturated compounds fall into a number of homologous series: (1) those containing one double bond (ethylene and its homologs) and having the formula C_nH_{2n} are known as olefins or alkenes; (2) those containing one triple bond (acetylene and its homologs) are called acetylenes or alkynes

TYPICAL ALIPHATIC HYDROCARBONS

$CH_3{-}CH_3$ Ethane

$CH_2{=}CH_2$ Ethylene

$CH{\equiv}CH$ Acetylene

$CH_2{=}C{=}CH{-}CH_3$ 1,2-Butadiene

$$\begin{array}{l} CH_2{=}\underset{\displaystyle CH_3}{\underset{|}{C}}{-}CH{=}CH_2 \end{array}$$

2-Methyl-1,3-butadiene (isoprene)

$CH_2{=}CH{-}CH_2{-}CH{=}CH_2$ 1,4-Pentadiene

$CH{\equiv}C{-}C{\equiv}C{-}CH_3$ 1,3-Pentadiyne

$$CH_3{-}\underset{\displaystyle CH_3}{\underset{|}{C}}{=}CH{-}CH{=}CH{-}\underset{\displaystyle CH_3}{\underset{|}{C}}{=}CH{-}CH_3$$

2,6-Dimethyl-2,4,6-octatriene (alloocimene)

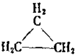
$CH_2{=}CH{-}C{\equiv}CH$ 1-Buten-3-yne

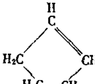
$CH_2{=}CH{-}CH{=}CH{-}C{\equiv}CH$ 1,3-Hexadien-5-yne

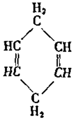
and have the formula C_nH_{2n-2}; (3) those having two double bonds (allene, 1,3-butadiene, and 1,4-pentadiene represent three types) are diolefins or alkadienes and also have the formula C_nH_{2n-2}; (4) those having a larger number of double or triple bonds or of both double and triple bonds are named in analogous fashion as alkatrienes, alkatetraenes, alkadiynes, alkenynes, and alkadienynes. *See* ALIPHATIC HYDROCARBON.

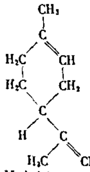
Alicyclic hydrocarbons. These nonaromatic cyclic (ring) compounds fall into a larger number of classes than do the aliphatic hydrocarbons because the rings may be of various sizes, they may be saturated or unsaturated, and the individual members may contain one or more rings. The saturated monocyclic compounds, C_nH_{2n}, known as cycloparaffins or cycloalkanes, include a large number of homologous series of which those containing five or six carbon atoms in the ring (cyclopentanes and cyclohexanes) are the most stable and common. The unsaturated monocyclic hydrocarbons include the cycloolefins, or cycloalkenes (C_nH_{2n-2}), and

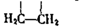
TYPICAL ALICYCLIC HYDROCARBONS

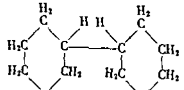

Cyclopropane

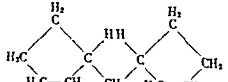

Cyclopentene

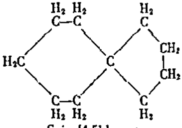

1,4-Cyclohexadiene

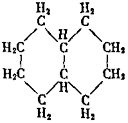

1-Methyl 4-isopropenylcyclohexene (limonene)

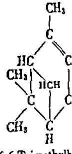

Methylenecyclobutane


Bicyclohexyl


Dicyclopentylmethane


Spiro[4.5]decane


Decahydronaphthalene


2,6,6-Trimethylbicyclo[3.1.1]-2-heptene (α-pinene or 2-pinene)

cyclodiolefins, or cycloalkadienes (C_nH_{2n-4}). There are very few examples of cycloacetylenes, or cycloalkynes, since a ring containing a triple bond would be under strain and therefore unstable.

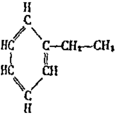
The word olefin is sometimes used as a generic term to include not only alkenes but also cycloalkenes and hydrocarbons containing more than one ethylenic double bond.

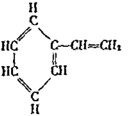
The rings in polycyclic alicyclic hydrocarbons may be connected in a variety of ways. These are typified by bicyclohexyl, dicyclopentylmethane, spiro[4.5]decane, decahydronaphthalene, and 2-pinene.

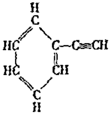
The saturated alicyclic hydrocarbons are sometimes called naphthenes, particularly by petroleum chemists. The word naphthene is related to the fact that cyclopentane and cyclohexane homologs have been isolated from the naphtha fraction of petroleum. *See* ALICYCLIC HYDROCARBON.

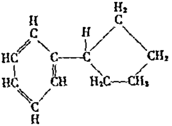
Aromatic hydrocarbons. These compounds contain at least one 6-membered benzene ring, incorporating what appear to be three conjugated double bonds. However, none of these bonds actually have the olefinic character associated with alkenes and cycloalkenes. The aromatic hydrocarbons fall into a number of types: (1) benzene, alkylbenzenes (ethylbenzene), alkenylbenzenes (styrene), and alkynylbenzenes (phenylacetylene); (2) cycloalkylbenzenes (cyclopentylbenzene) and cycloalkenylbenzenes (phenylcyclohexenes); (3) fused ring polynuclear hydrocarbons (indene, indan, naphthalene, tetrahydronaphthalene, fluorene, anthracene); (4) polynuclear aromatic hydrocarbons having directly united rings (biphenyl, binaphthyl) or rings united through aliphatic carbon (diphenylmethane, 1,2-diphenylethylene, diphenylacetylene). *See* AROMATIC HYDROCARBON.

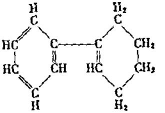
TYPICAL AROMATIC HYDROCARBONS

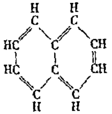

Ethylbenzene

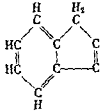

Vinylbenzene (styrene)

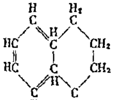

Phenylacetylene

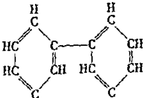

Cyclopentylbenzene

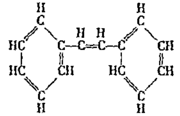

1-Phenylcyclohexene


Naphthalene


Indene


1,2,3,4-Tetrahydronaphthalene


Biphenyl


1,2-Diphenylethylene

Natural sources of hydrocarbons. Of these sources, the largest are natural gas and petroleum. The composition of natural gas varies greatly depending on the area from which it is obtained. The methane content usually ranges from about 50 to about 90%; ethane, from about 5 to about 20%; propane, from about 3 to about 18%; and butanes from about 1 to about 7%.

Petroleum is a complex mixture of liquid and solid hydrocarbons, its composition also varying with the source. Its chief components are paraffinic hydrocarbons, saturated five- and six-carbon atom alicyclic hydrocarbons, and aromatic hydrocarbons. Olefins are usually absent. Petroleum is distilled into a number of commercial fractions: gasoline, boiling at about 20–200°C; kerosine, 175–275°C; and heating oils, 250–400°C. Refining of higher boiling fractions by treatment with acids or clays and by crystallization yields lubricating oils, white mineral oils, petrolatums (petroleum jelly), and paraffin wax, composed predominantly of straight-chain alkanes. See PETROLEUM PROCESSING.

Some normal paraffins have been isolated from vegetable products. For example, *n*-heptane, used as a standard fuel in determination of the knock rating (octane number) of gasoline, was formerly obtained by distillation and chemical treatment of the oil obtained from the resin of the Jeffrey pine; it is now prepared by distillation from straight-run gasoline.

Hydrocarbons are obtained by the destructive distillation or carbonization of coal (heating in the absence of air) at 350–1000°C, producing coal gas and coal tar. Coal gas contains methane together with smaller amounts of ethane, ethylene, benzene, toluene, cyclopentadiene, naphthalene, and nonhydrocarbon products such as hydrogen, ammonia, carbon monoxide, carbon dioxide, hydrogen sulfide, hydrogen cyanide, cyanogen, and nitric oxide. The normally liquid or solid hydrocarbons in the gas (namely the cyclopentadiene, benzene, toluene, and naphthalene) can be recovered. The gas is further treated to remove toxic components and then used as a domestic fuel.

The volatile oils obtained from certain plants (caraway or dill), trees (pine), and citrus fruits (lemon or orange) contain members of a large group of hydrocarbons known as terpenes, having the formula $C_{10}H_{16}$. These exist both as aliphatic compounds (alkatrienes such as alloocimene, or 2,6-dimethyl-2,4,6-octatriene) or more usually, as alicyclic compounds (limonene or 1-methyl-4-

propenylcyclohexene, α-pinene, or 2,6,6-trimethylbicyclo[3.1.1]-2-heptene). Related hydrocarbons are the sesquiterpenes, $C_{15}H_{24}$; diterpenes, $C_{20}H_{32}$; triterpenes, $C_{30}H_{48}$; and polyterpenes, $C_{10n}H_{16n}$.

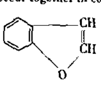
Rubber, which like the terpenes is a polymer of isoprene, may be considered an open-chain polyterpene. The rubber molecule contains one double bond per isoprene unit. *See* TERPENE.

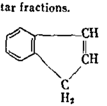
[L.S.]

Hydrocarbon resin

Brittle or gummy materials prepared by the polymerization of several unsaturated constituents of coal tar, rosin, or petroleum. The hydrocarbon resins are inexpensive and find many uses in rubber and asphalt formulations and in coating and calking compositions. Brief descriptions will be given of the coumarone-indene resins, the petroleum resins, and the polyterpene resins.

Coumarone-indene resins. Coumarone and indene occur together in coal-tar fractions.


Coumarone

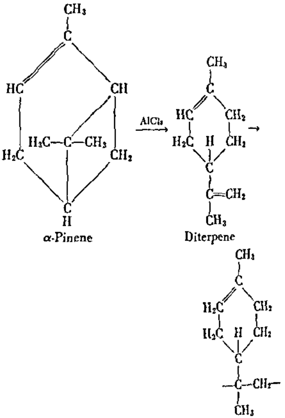

Indene

Indene usually predominates, and the mixture may contain small amounts of cyclopentadiene and styrene. The mixture polymerizes readily in the presence of sulfuric acid to yield soft to hard, brittle polymers. The molecular weights vary from about 500 to 4000, and the colors of the products range from yellow to dark brown. The soft products are used as plasticizers and tackifiers in calking formulations and rubber compositions. The hard products are used to stiffen and strengthen synthetic rubber, and they also are used in floor coverings, roofing materials, varnishes, and metal paints and in combination with other resins.

Indene without coumarone can be obtained from petroleum naphtha, and it polymerizes to yield products which are generally similar to the coumarone-indene resins.

Petroleum resins. Mixtures of polymerizable dienes and unsaturated compounds are obtained from gasoline refining operations. The properties and uses of the resins are generally similar to those of the coumarone-indene resins. The resins are recommended for use in paper coatings and hot-melt adhesive formulations. Aqueous suspensions can be employed in making water-based paint formulations.

Polyterpenes. The α- and β-pinene obtained from turpentine polymerize in the presence of acid catalysts with the formation of soft to hard polymers of molecular weights up to about 2000. The hard products have softening points up to about 100°C. The rearrangement and polymerization of α-pinene are shown in the following equation:



The properties and uses of the polyterpenes are generally similar to those of the other hydrocarbon resins described above. They are especially recommended for use in oil paints and adhesive formulations. *See* POLYMERIZATION. [J.A.M.; L.M.H.]

Hydrodynamics

That branch of continuum mechanics which treats of the laws of motion of an incompressible fluid and of the interactions of the fluid with its boundaries. Partly because of the age-old interest in systems of water works and water-borne vehicles, and partly because of scientific curiosity, the subject has achieved a high state of both practical and analytical development. The present article presents a brief rational sketch of the state of this subject. For more detail on specific parts of the subject *see* FLOW OF FLUIDS; FLUID MECHANICS; FLUID-FLOW PRINCIPLES; FLUID-FLOW PROPERTIES.

Flow field. Fluids have many physical properties, including, besides thermodynamic and electromagnetic properties, the dynamical ones of density, viscosity, adhesion, and cohesion. Only the dynamical properties will be considered in this article, and furthermore, because capillary effects will not be treated, the last two properties will be of interest only insofar as they affect the inception of cavitation.

A flow field can be studied from either of two points of view, the Lagrangian, in which one is concerned with the history of each fluid particle, and the Eulerian, in which interest is focused

rather, on the velocity vectors associated with each point. Although the Lagrangian method is convenient for demonstrating certain kinematical properties of a flow field, the alternative approach will be used, because it has been found, on the whole, to be simpler and more powerful. *See* EULER'S EQUATIONS OF MOTION; LAGRANGE'S EQUATIONS.

A fruitful concept in hydrodynamics is that of the ideal or inviscid fluid. In such a fluid, across any element of area, the fluid on one side exerts a normal stress or pressure on the fluid on the other side, the magnitude of which is independent of the orientation of the element. Thus there is a scalar pressure field associated with the vector velocity field. A principal problem of hydrodynamics is to *determine these fields corresponding to prescribed* boundary conditions.

When viscosity is taken into account, the normal stresses across an area element are not, in general, independent of orientation, and in addition, tangential stresses are present. Instead of a scalar pressure field one must consider now a tensor stress field. These stresses, or their integrals, must be determined in order to solve the important problem of obtaining the force and moment acting on a body moving through the fluid. *See* NEWTONIAN FLUID; VISCOSITY OF LIQUIDS.

Hydrokinematics. For the present purpose, the molecular structure of a fluid can be ignored and it can be treated as a continuum. Thus various properties of the fluid will be considered as continuous functions of space and time. If the position of a point is represented in terms of its coordinates x, y, z with respect to a rectangular cartesian coordinate system, the velocity vector field may be expressed in the form $\mathbf{U}(x,y,z,t)$, with components $u(x,y,z,t)$, $v(x,y,z,t)$, $w(x,y,z,t)$ and magnitude

$$V = \sqrt{u^2 + v^2 + w^2}$$

Similarly the mass density of the fluid (mass per unit volume), probably its most important property, may be shown as the function $\rho(x,y,z,t)$.

A basic relation, which expresses the law of conservation of mass, is the equation of continuity which, in regions containing no fluid sources or sinks, may be written in the form

$$\operatorname{div}(\rho\mathbf{U}) = \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = -\frac{\partial \rho}{\partial t}$$

If ρ is constant, as is assumed hereafter, the equation becomes

$$\operatorname{div}\mathbf{U} = 0$$

From the latter equation can be deduced the existence of two families of stream surfaces

$$\psi(x,y,z) = a \qquad \chi(x,y,z) = b$$

defined by constant values of a and b, the mutual *intersections of which define the* streamlines of the flow pattern, that is, lines tangent at a given instant to the velocity vectors (*see* STREAMLINE FLOW). According to this definition, the differential equations of a streamline are

$$\frac{dx}{u(x,y,z,t_0)} = \frac{dy}{v(x,y,z,t_0)} = \frac{dz}{w(x,y,z,t_0)}$$

This should be distinguished from the equations of a path line

$$\frac{dx}{u(x,y,z,t)} = \frac{dy}{v(x,y,z,t)} = \frac{dz}{w(x,y,z,t)}$$

which gives the path followed by a fluid particle. If the flow is steady (time-independent) the two sets of lines are coincident. In any case they are tangent at the location of a particle.

If the stream functions ψ and χ have been determined for a flow problem, the velocity vector would be given by the vector cross product

$$\mathbf{U} = \operatorname{grad}\psi \times \operatorname{grad}\chi$$

These stream functions also have the significant *property that the rate of flow Q through a stream* channel bounded by the four surfaces ψ_1, χ_1, ψ_2, χ_2 is (Fig. 1)

$$Q = (\psi_2 - \psi_1)(\chi_2 - \chi_1)$$

In two important special cases the stream functions reduce to a single one. For two-dimensional flow there is the Lagrange stream function $\psi(x,y)$ in terms of which the velocity components are

$$u = \frac{\partial\psi}{\partial y} \qquad v = -\frac{\partial\psi}{\partial x}$$

Constant values of ψ give the streamlines, and the rate of flow q between a pair of streamlines ψ_1 and ψ_2 is

$$q = \psi_2 - \psi_1$$

The other case is that of axisymmetric flow, for which there is the Stokes stream function $\psi(r,z)$, where the z axis is coincident with the axis of symmetry and r denotes distance perpendicular to it (*see* STOKES STREAM FUNCTION). The velocity components are now

$$u = -\frac{1}{r}\frac{\partial\psi}{\partial z} \qquad w = \frac{1}{r}\frac{\partial\psi}{\partial r}$$

and the rate of flow between two stream surfaces

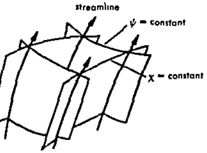


Fig. 1. Stream channel formed by stream surfaces

ψ_1 and ψ_2 is

$$Q = 2\pi(\psi_2 - \psi_1)$$

A useful concept in hydrodynamics is that of sources or sinks at various points of a fluid at which fluid is entering or leaving a region (*see* SINK FLOW; SOURCE FLOW). If $4\pi M$ (where M is called the source strength) denotes the volume rate of entry at a source (or sink, if negative), it is seen, by considering the discharge through a sphere of radius r with its center at the source, that the fluid velocity due to the source is radial and of magnitude $V = M/r^2$. If there are many sources in a region, Gauss' theorem states that the flux of fluid through the surface of this region is the product of the sum of the source strength by 4π (*see* GAUSS' THEOREM). If the sources are distributed continuously through a region with a density m per unit volume, so that $m d\tau$ is the strength of the sources in a volume element $d\tau$, it is necessary to generalize the equation of continuity to become

$$\text{div } \mathbf{U} = 4\pi m$$

Analysis of the behavior of a fluid element in a stream shows that it is rotating at an angular velocity $\omega/2$, where

$$\omega = \text{curl } \mathbf{U}$$

is called the vorticity vector (*see* VORTEX). This vector is closely related to the circulation, defined as the line integral around a closed curve of the tangential component of the velocity,

$$\Gamma = \oint \mathbf{U} \cdot d\mathbf{s}$$

By application of Stokes' theorem it is readily shown that the circulation is also given by the surface integral of the vorticity over a surface bounded by the curve

$$\Gamma = \int_s \omega \cdot \mathbf{n}\, dS$$

where $\mathbf{n}$ is the unit normal vector at a point of the surface, positive in the sense of advance of a right-hand screw relative to the positive sense of describing the closed curve.

Lines tangent to the vorticity vector at every point are called vortex filaments and a group of vortex lines can form a vortex tube. These have the property that a vortex tube can begin or end only at a boundary, unless it is in the form of a ring. Furthermore, the circulation is constant at all sections of a vortex tube and is called its strength. Two additional laws of vortex motion, for an inviscid fluid on which only conservative forces such as gravitational attraction are acting, are (1) vortex filaments are always composed of the same fluid particles, and (2) the strength of a vortex tube is constant with respect to time.

A straight-line vortex filament of strength Γ induces a velocity $V = \Gamma/r$ tangent to a circle of radius r with center on the line in the plane perpendicular to it.

A flow field can be determined when its distributions of sources and vortices are prescribed. The problem is to determine the velocity $\mathbf{U}$ from the equations

$$\text{div } \mathbf{U} = 4\pi m \qquad \text{curl } \mathbf{U} = \omega$$

This is accomplished by putting $\mathbf{U} = \mathbf{U}_1 + \mathbf{U}_2$, where

$$\begin{aligned} \text{div } \mathbf{U}_1 &= 4\pi m \qquad & \text{curl } \mathbf{U}_1 &= 0 \\ \text{div } \mathbf{U}_2 &= 0 & \text{curl } \mathbf{U}_2 &= \omega \end{aligned}$$

The velocity field $\mathbf{U}_1$ is said to be irrotational, $\mathbf{U}_2$ to be solenoidal. Then

$$\mathbf{U}_1 = \text{grad } \phi$$

$$\phi = -\int \frac{m(\xi,\eta,\zeta)\, d\xi\, d\eta\, d\zeta}{[(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2]^{1/2}}$$

$$\mathbf{U}_2 = \text{curl } \mathbf{A}$$

$$\mathbf{A} = \frac{1}{4\pi} \int \frac{\omega(\xi,\eta,\zeta)\, d\xi\, d\eta\, d\zeta}{[(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2]^{1/2}}$$

The functions ϕ and $\mathbf{A}$ are called the scalar and vector potentials.

Hydrokinetics. A fluid element is acted upon by body forces (forces per unit mass) such as gravitational attraction, and stresses on its surface (forces per unit area). Let $\mathbf{B}$, with components B_x, B_y, B_z, denote the body force. The stresses can be represented by the symmetric tensor τ_{ij}, where the index i indicates that the stress is acting on a plane with positive normal in the direction of increasing x, y, or z according as $i = 1, 2,$ or 3; and $j = 1, 2,$ or 3 indicates the x, y, or z component of the stress (Fig. 2). In an inviscid fluid this stress tensor reduces to the scalar pressure field.

The force $\mathbf{F}$ and moment $\mathbf{M}$ on a body in a fluid then assume the following forms for an inviscid fluid,

$$\begin{aligned} \mathbf{F} &= -\int p\mathbf{n}\, dS + \rho \int \mathbf{B}\, d\tau \\ \mathbf{M} &= -\int p\mathbf{r} \times \mathbf{n}\, dS + \rho \int \mathbf{r} \times \mathbf{B}\, d\tau \end{aligned}$$

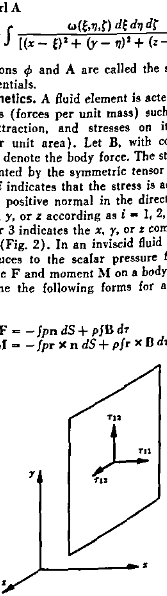


Fig. 2. Normal and tangential stresses.

for a viscous fluid,

$$X = \sum_{i=1}^{3} \int \tau_{1i} n_i \; dS + \rho \int B_x \; d\tau$$
$$Y = \sum_i \int \tau_{2i} n_i \; dS + \rho \int B_y \; d\tau$$
$$Z = \sum_i \int \tau_{3i} n_i \; dS + \rho \int B_z \; d\tau$$
$$L = \sum_i \int (y\tau_{3i} - z\tau_{2i}) n_i \; dS + \rho \int (yB_z - zB_y) \; d\tau$$
$$M = \sum_i \int (z\tau_{1i} - x\tau_{3i}) n_i \; dS + \rho \int (zB_x - xB_z) \; d\tau$$
$$N = \sum_i \int (x\tau_{2i} - y\tau_{1i}) n_i \; dS + \rho \int (xB_y - yB_x) \; d\tau$$

Application of Newton's laws of motion to an element of an inviscid fluid leads to the Euler equations

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} = B_x - \frac{1}{\rho}\frac{\partial p}{\partial x}$$
$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} = B_y - \frac{1}{\rho}\frac{\partial p}{\partial y}$$
$$\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} = B_z - \frac{1}{\rho}\frac{\partial p}{\partial z}$$

which may also be written in the vector form

$$\frac{\partial \mathbf{U}}{\partial t} + \boldsymbol{\omega} \times \mathbf{U} + \tfrac{1}{2}\,\mathrm{grad}\,(V^2) = \mathbf{B} - \frac{1}{\rho}\,\mathrm{grad}\,p$$

The latter form is useful when it is desired to transform the equations of motion to other coordinate systems.

It will be supposed that the body force is a conservative one so that it can be expressed in the form $\mathbf{B} = \mathrm{grad}\,\Omega$ where Ω is a scalar potential function. For example, if z is taken positive upwards at a point on the surface of the earth, the scalar potential for gravitational attraction is $\Omega = -gz$, where g is the acceleration of gravity.

Probably the most frequently applied result of hydrodynamics is the Bernoulli equation which can be derived as a first integral of the Euler equations. This assumes different forms for various assumed conditions:

1. If the flow is steady, then for points along a streamline or along a vortex line

$$p + \tfrac{1}{2}\rho V^2 + \rho\Omega = \text{constant}$$

2. In a region in which the flow is irrotational the equation is

$$p + \tfrac{1}{2}\rho V^2 + \rho\Omega + \frac{\partial \phi}{\partial t} = \text{constant}$$

3. If the flow is steady and the streamlines and vortex lines are parallel, the equation is

$$p + \tfrac{1}{2}\rho V^2 + \rho\Omega = \text{constant}$$

4. If the coordinate axes are in motion, the origin having the velocity components U, V, W, and the coordinate system the angular velocity $\mathbf{\Lambda}$, then

$$p + \tfrac{1}{2}\rho[(u-U)^2 + (v-V)^2 + (w-W)^2] + \rho\Omega + \mathbf{\Lambda} \cdot \mathbf{r} \times \mathbf{U} + \frac{\partial \phi}{\partial t} = \text{constant}$$

in which $\mathbf{\Lambda} \cdot \mathbf{r} \times \mathbf{U}$ denotes the triple scalar product of the indicated vectors, and $\mathbf{U}$ and its components u, v, w are velocities of the fluid. *See* BERNOULLI'S THEOREM.

If a body is immersed in a steady stream of velocity and pressure V_∞ and p_∞ at infinity, and if V and p denote the velocity and pressure at a point on the body, the Bernoulli equation may be written in the dimensionless form

$$\frac{p - p_\infty}{\tfrac{1}{2}\rho V_\infty^2} = 1 - \left(\frac{V}{V_\infty}\right)^2$$

At the point of maximum velocity on the body the pressure has a minimum value, denoted by p_m. Because ratio $(p_m - p_\infty)/(\tfrac{1}{2}\rho V^2)$ is a constant for the body, the value of p_m can be reduced either by reducing the ambient pressure p_∞ (as can be done in a variable-pressure water tunnel) or by increasing V_∞. If by either means p_m is reduced to the vapor pressure of a liquid p_v, and the liquid begins to vaporize, the liquid is said to undergo cavitation, and

$$\sigma_v = \frac{p_\infty - p_v}{\tfrac{1}{2}\rho V_\infty^2}$$

is called the vapor-pressure cavitation number. The phenomenon of cavitation is of great technical importance in the design of ship propellers, turbines, and other hydraulic structures because of the erosion caused by the collapsing cavitation bubbles (*see* CAVITATION). The inception pressure may be greater than the vapor pressure if a considerable amount of entrained air is present; on the other hand, a specimen of liquid may be able to withstand tensions of thousands of atmospheres if special care has been taken to remove gas nuclei from it.

When viscosity is taken into account, the equations of motion become the Navier-Stokes equations

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} = B_x - \frac{1}{\rho}\frac{\partial p}{\partial x} + \nu\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right)$$
$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} = B_y - \frac{1}{\rho}\frac{\partial p}{\partial y} + \nu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}\right)$$
$$\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} = B_z - \frac{1}{\rho}\frac{\partial p}{\partial z} + \nu\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}\right)$$

which may also be written in the vector form

$$\frac{\partial \mathbf{U}}{\partial t} + \boldsymbol{\omega} \times \mathbf{U} + \tfrac{1}{2}\,\mathrm{grad}\,(V^2) = \mathbf{B} - \frac{1}{\rho}\,\mathrm{grad}\,p - \nu\,\mathrm{curl}\,\boldsymbol{\omega}$$

in which ν is the kinematic viscosity (*see* NAVIER-STOKES EQUATIONS). These and the equation of continuity give four equations for solving for u, v, w, and p. If the equations can be solved, the stresses can then be obtained from the fundamental relations between stresses and rates of strain

$$\tau_{11} = -p + 2\mu\frac{\partial u}{\partial x} \qquad \tau_{22} = -p + 2\mu\frac{\partial v}{\partial x}$$

$$\tau_{33} = -p + 2\mu\frac{\partial w}{\partial x} \qquad \tau_{23} = \mu\left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}\right)$$

$$\tau_{31} = \mu\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right) \qquad \tau_{12} = \mu\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)$$

where μ is the coefficient of dynamic viscosity.

Two immediate and important consequences of the Navier-Stokes equations are (1) the circulation in a closed circuit moving with the fluid diminishes at a time rate given by $\nu\oint(\mathrm{curl}\,\boldsymbol{\omega})\cdot d\mathbf{s}$, and (2) the energy per unit volume of a fluid diminishes at a time rate given by

$$\mu\left[2\left(\frac{\partial u}{\partial x}\right)^2 + 2\left(\frac{\partial v}{\partial y}\right)^2 + 2\left(\frac{\partial w}{\partial z}\right)^2 + \left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}\right)^2 + \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)^2\right]$$

These furnish a measure of the effect of viscosity on the rates of dissipation of both vorticity and energy.

Irrotational flow. In irrotational flow the vorticity is zero and the velocity is expressible as the gradient of a scalar potential $\mathbf{U} = \mathrm{grad}\,\phi$. The equation of continuity in rectangular coordinates then assumes the form of Laplace's equation

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = 0$$

When the flow is irrotational, the terms due to viscosity in the Navier-Stokes equations vanish so that they become formally identical with the Euler equations. Thus, in either viscous or inviscid irrotational flow, if Laplace's equation has been solved for the velocity potential, the pressure field can be obtained from the Bernoulli equation.

Elementary solutions of Laplace's equation are

$$\phi = V(lx + my + nz)$$

for uniform flow in the direction of a line having direction cosines l, m, n;

$$\phi = -\frac{M}{(x^2 + y^2 + z^2)^{1/2}}$$

for a source of strength M at the origin; and

$$\phi = -\frac{x\Delta}{(x^2 + y^2 + z^2)^{3/2}}$$

for a doublet of strength Δ oriented along the x axis. A doublet may be obtained by letting a source and sink of equal strength approach each other (*see* DOUBLET FLOW). The product of the source strength by the distance between them is held constant (Fig. 3). By combining these elementary solutions many interesting flow patterns can be obtained; for example

$$\phi = Vx\left[1 + \frac{a^3}{2(x^2 + y^2 + z^2)^{3/2}}\right]$$

formed by adding the potentials of a uniform stream and a doublet, gives the flow about a sphere. Bodies obtained as stream surfaces by trying various combinations of sources, sinks, and doublets in a uniform stream are called Rankine bodies. A necessary condition for such bodies to be closed is that the algebraic sum of the strengths of the sources and sinks be zero.

Methods of solving the direct problem, of finding the flow subject to prescribed boundary conditions, have also been devised. These methods may be classified in five categories.

Separation of variables. This reduces Laplace's equation to several ordinary differential equations, the solutions of which are obtained as sets of orthogonal functions. Combinations of these are then found which satisfy the boundary conditions.

Method of images. This method is suitable when the boundaries are planes, spheres, or circular cylinders. This technique has recently reached its culmination in the discovery of so-called sphere and circle theorems which immediately give the modification of the flow when a sphere or circle is introduced into a preexisting flow pattern.

Method of integral equations. Two important classes of problems, the Neumann problem, in which the normal component of the velocity is prescribed on the boundary, and the Dirichlet problem, in which the values of the potential are given on the boundary, can be formulated as

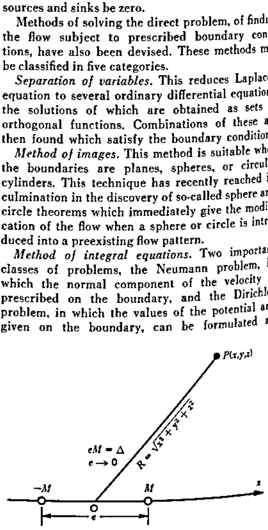


Fig. 3. Definition sketch of doublet.

Fredholm integral equations. Although it is tedious to solve these equations numerically by hand, programs are available for solving them on moderate- or high-speed automatic computers.

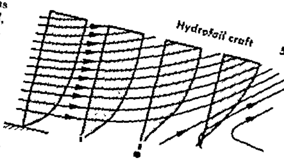
Solution by relaxation. This is a numerical method in which the flow region is divided into a fine network, at the corners of which the values of ϕ are assumed as an initial approximation. The finite-difference form of Laplace's equation and the boundary conditions are then used to adjust the assumed values by trial and error. This method is also tedious.

Conformal mapping. The theory of functions of a complex variable furnishes a remarkably powerful tool for solving two-dimensional irrotational flow problems. Riemann's mapping theorem gives assurance that boundaries of regions can be mapped into circles by means of analytic functions. If the proper function is found, then it also transforms the flow about these boundaries into the much simpler problem of the flow about the circles, which can be solved. The difficulty in conformal mapping lies in finding the appropriate mapping functions. Many flow problems with simple boundaries can be solved by applying the properties of elementary functions of complex variables. For arbitrary shapes, several numerical procedures have been developed.

Viscous flow. Because the viscosity of the most common fluids, air and water, is very small, it is an excellent approximation to treat their flows as inviscid except at very low Reynolds numbers ($Vl/\nu < 10$) and in the neighborhood of the boundaries. Because the velocity of a fluid at a wall is zero relative to that of the boundary (nonslip condition), a so-called boundary layer in which viscous effects are important is present, within which appreciable shear stresses occur. Outside of this boundary layer the flow may be treated as inviscid and to a good approximation, the pressure across the boundary layer may be assumed to be that in the inviscid flow at the outer edge of the boundary layer.

A simplified form of the Navier-Stokes equations, called the boundary-layer equations, yields laminar-flow solutions for the velocity profiles. It is known, however, that boundary-layer flows are often turbulent. A theory of the stability of the laminar boundary layer has been developed and current research is shedding considerable light on the processes whereby a disturbed laminar flow eventually breaks down into the typical random motions of turbulence, but the mechanisms are not yet clear.

Although turbulent flow is believed to be governed by the Navier-Stokes equations, it has not yet been possible to derive any turbulent-flow solutions from these equations. The difficulty lies in the fact that the Reynolds equations for turbulent flow, which are derived from the Navier-Stokes equations by an averaging process, introduce new unknowns for which additional relations are required. The well-developed theory of homogeneous isotropic turbulence, and the body of accumulated experimental results and hypotheses concerning turbulence in shear flows, have not yet supplied these missing relations.



Fi

paration.

A phenomenon of real flow that is of great practical concern is that of separation of flow. This occurs in a boundary layer when the fluid in the neighborhood of a wall is brought to rest and caused to reverse its direction by the action of an adverse pressure gradient, that is, increasing pressures in the downstream direction (see Fig. 4). When the boundary layer is turbulent, separation occurs farther downstream (or not at all) than when it is laminar, because the momentum of the layers near the wall is reinforced by turbulent interchange with layers of higher mean velocity due to the random motions. See BOUNDARY-LAYER FLOW; LAMINAR FLOW; TURBULENT FLOW. [L.LA.]

Bibliography: S. Goldstein (ed.), *Modern Developments in Fluid Dynamics*, 1938; H. Lamb, *Hydrodynamics*, 6th ed., 1945; H. Rouse (ed.), *Advanced Mechanics of Fluids*, 1959.

Hydroelectric generator

An electric rotating machine which transforms mechanical power from a hydraulic turbine, or water wheel, into electric power. Hydroelectric generators may have horizontal or vertical shafts, depending upon the turbine. The most common type in large ratings is the vertical-shaft, synchronous generator. For other generators see GENERATOR, ELECTRIC; see also ELECTRIC POWER GENERATION. [L.T.R.]

Hydrofoil craft

Marine craft whose hulls are borne clear of the water, dynamically sustained by a system of sub-surface hydrofoils. At low speeds such craft operate as though they possessed conventional displacement hulls. As the speed increases, the foils support an increasing proportion of the craft weight until the hull finally rises and travels clear of the water. The lift principle employed by these foils is identical with that of the airfoil except for additional complications resulting from the proximity and disturbed condition of the sea surface, cavitation, and ventilation. See AIRFOIL.

The principal advantages of this craft are its efficiency at high speed (achieved by elimination

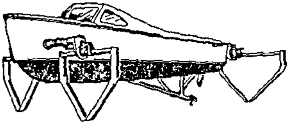


Fig 1. Surface-piercing V-foil craft. (Baker Mfg. Co.)

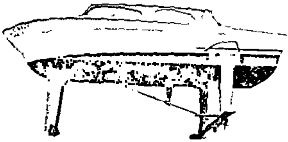


Fig 2. Submerged horizontal-foil craft. (Gibbs and Cox, Inc.)

of the skin friction and wave drag normally associated with the hull), and its relative freedom of motion in a seaway. For a discussion of skin friction and wave drag, *see* Ship propulsion.

Although a successful hydrofoil craft was built by Alexander Graham Bell in 1918, only since 1952 have such craft been employed in extensive commercial operation. The range in displacement has been from 0.1 to 80 tons and in maximum speed [illegible]

[illegible] (Fig. 1) or fully submerged horizontal foils (Fig. 2). Many variants of these configurations have been tested, including ladders of surface-piercing foils, combinations of foils and planing surfaces, and sweep-back design.

The high speed and sustained hull dictate unusual propulsion characteristics. Conventional propellers must be deeply submerged to avoid cavitation, and thus require an angled or geared drive to [illegible]

[illegible] propeller is one whose suction face is enveloped by a vapor cavity. For details, *see* Propeller, marine.

The attainment of adequate sea stability is a primary design problem. V-foil systems have inherent lateral and longitudinal trim stability, and craft elevation [illegible]

[illegible] the direction in which the sea is moving, however [illegible]

[illegible] the angle of attack of the foil system in response to changes in pitch, heel, and elevation.

Other problems in hydrofoil craft arise from the sucking of air down the supporting struts into the low-pressure face of the foil, cavitation, and vibration of the foils. The former problem can be partially alleviated by the use of fins ("fences") located on the struts near the surface of the water.

[J.P.CR.]

Bibliography: T. M. Buermann, P. Leehey and J. J. Stilwell, An appraisal of hydrofoil supported craft, *Soc. Naval Architects Marine Engs. Trans.* 61:242–279, 1953.

Hydroformylation

The synthesis of aliphatic hydrocarbons along with some oxygenated compounds from mixtures of carbon monoxide and hydrogen over an alkalized iron catalyst is generally known as the Fischer-Tropsch reaction. Hydrocarbons are the principal products of this reaction, but by the use of selected catalysts and operating conditions, the yield of oxygenated derivatives can be increased. In the related Oxo process, it is feasible to obtain the exclusive production of oxygenated aliphatic compounds. *See* Fischer-Tropsch reaction.

In the Oxo process, the term hydroformylation applies to those reactions brought about by treatment of olefins with a mixture of hydrogen and carbon monoxide in the presence of a cobalt catalyst. Thus, the conversion of an olefin to an aldehyde containing an additional carbon atom occurs by the addition of the formyl (CHO) group to a double bond.

$$\mathrm{R{-}CH{=}CH_2 + CO + H_2} \begin{cases} \nearrow \ \mathrm{R{-}(CH_2)_2{-}CHO} \\ \searrow \ \mathrm{R{-}\underset{\displaystyle CHO}{\underset{|}{CH}}{-}CH_3} \end{cases}$$

By treatment with hydrogen in an integrated system, the aldehydes are converted to alcohols. When propylene, heptenes, and nonenes are processed by these reactions, normal and isobutyl alcohols, isooctyl alcohol, and primary decyl alcohols, respectively, are obtained. The production of Oxo alcohols, many of which enter into the manufacture of surfactants, plasticizers, and lubricants, is a major industry.

Reactants. Synthesis gas and olefins are the reactants. The ratio of H_2 to CO is usually 1:1. Although equal molecular quantities of synthesis gas and olefin are consumed, an excess of synthesis gas is normally used. Hydrogen, which is required for the second step to reduce the aldehydes, can be obtained as a by-product from refinery operations or from synthesis gas.

$$\mathrm{CO + H_2O \rightarrow CO_2 + H_2}$$

Olefins of many types ranging from ethylene to hexadecene have been found to be reactive. Compounds such as divinyl ether, ethyl oleate, and cyclopentadiene have also been used. Mixtures of olefins react as readily as pure compounds, and the

presence of saturated compounds does not interfere with the reaction.

Cobalt catalysts are universally used for the Oxo reaction. For operations above 150°C and about 3000 psi, almost any salt of cobalt can be used. At lower temperatures, dicobalt octacarbonyl or a Raney cobalt must be employed.

Mechanism. As indicated above, the hydroformylation reaction consists of the addition of H and CHO to the carbon atoms joined by a double bond. With a straight-chain unsymmetrical olefin, addition can lead to one or the other of two compounds or to a mixture of both depending on where the formyl group is added. When the double bond is in the terminal position, the product is usually a mixture of aldehydes in which straight- and branched-chain isomers occur in almost equal proportions. When the olefinic reactant has the double bond in other than the terminal position, the products are nevertheless almost identical to those obtained from the isomer having a terminal double bond. As a generalization it can be stated that a straight-chain olefin will yield about 60–40% normal alcohol and 40–60% α-branched-chain alcohol no matter where the double bond is located.

Kinetic studies indicate that the rate of hydroformylation (1) is a first-order reaction with respect to olefin; (2) is approximately proportional to the amount of cobalt present; (3) increases with increasing hydrogen pressure at constant carbon monoxide pressure; and (4) decreases with increasing carbon monoxide pressure at constant hydrogen pressure. Item 4 is the basis for the postulate that the first step involves the formation of an olefin-carbonyl complex and carbon monoxide by the reaction of dicobalt octacarbonyl with the olefin.

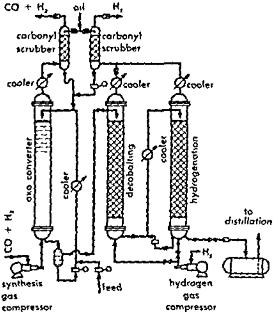
$$Co(CO)_8 + RCH{=}CH_2 \rightarrow Co(CO)_7RCH{=}CH_2 + CO$$

Decomposition of the complex by reaction with hydrogen or cobalt hydrocarbonyl is the suggested route to the formation of aldehyde and a precursor of dicobalt octacarbonyl.

The heat liberated in the hydroformylation reaction (about 30 kcal/mole) does not vary greatly with the structure of the olefin. In view of the magnitude of this exotherm, heat removal is an important factor in the design of a reactor.

Technical operations. The Oxo reaction is carried out in the liquid phase. With gaseous olefins, a suspension medium, such as an inert hydrocarbon that has a boiling range appreciably different from the aldehyde product, is used. Two liquid-phase processes have been used, namely, the slurry system and the fixed-bed process. Less operating difficulty is claimed for the latter system because the need for filtration of catalyst and solids handling is obviated.

In the fixed-bed process, shown schematically in the illustration, soluble cobalt salts of fatty acids or naphthenates are pumped with the olefin to the top of the first reactor and flow countercurrent to



Oxo process with fixed catalyst bed.

the synthesis gas. One type of fixed-bed catalyst consists of 2% metallic cobalt on a pumice carrier. Part of the cobalt is converted to carbonyl, leaves the reactor with the overhead product, and is replaced by cobalt salts in the feed. Unreacted synthesis gas leaving the top of the reactor is cooled, passed through a packed tower countercurrent to the olefin feed to remove cobalt carbonyl, and recycled to the reactor.

The second vessel is a decobalting converter in which cobalt carbonyl, dissolved in the product from the first reactor, is decomposed by treatment with hydrogen at about 200–220 atm and 120–150°C. The liquid enters at the top and flows countercurrent to the hydrogen. Metallic cobalt is deposited on the packing. Gas flows from the top of the decobalting unit to the carbonyl scrubber, and the liquid leaving the bottom is sent to the hydrogenation reactor for conversion to alcohol. Here, copper chromite is generally used as the catalyst. To ensure complete conversion of aldehydes to the

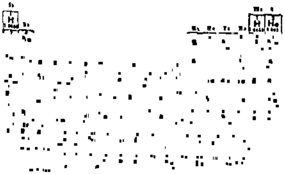
[F.H.C.]

Hydrogen

The first chemical element in the periodic system. Under ordinary conditions, it is a colorless, odorless, tasteless gas composed of diatomic molecules, H_2. The hydrogen atom, symbol H, consists of a nucleus of unit positive charge and a single electron. It has an atomic number of 1 and an atomic weight of 1.0080. The element is a major constituent of water and all organic matter, and is widely distributed not only on the earth but throughout the universe. There are three isotopes of hydrogen:

protium, mass 1, makes up 99.98% of the natural element; deuterium, mass 2, makes up about 0.02%; tritium, mass 3, occurs in extremely small amounts in nature but may be produced artificially by various nuclear reactions. *See* DEUTERIUM; ISOTOPE; TRITIUM.

Although hydrogen had been produced by earlier workers by the reaction of metals with acid, it was Henry Cavendish (1731–1810) who first distinguished it from other flammable gases.



Uses of hydrogen. A large fraction of the total hydrogen produced is used in the synthesis of ammonia. Ammonia plants are often built adjacent to petroleum refineries or coking plants to utilize the by-product hydrogen which might otherwise be wasted. Large quantities of hydrogen are also consumed in the catalytic hydrogenation of unsaturated liquid vegetable oils to make solid fats. Hydrogenation is used in the manufacture of organic chemicals, such as alcohols from esters and glycerides, amines from nitriles, and cycloparaffins from aromatic hydrocarbons. Methanol is synthesized on a commercial scale by the reaction of hydrogen with carbon monoxide. Hydrogen is also utilized for the production of hydrochloric acid by reaction with chlorine gas and in petroleum refining to remove sulfur and to hydrogenate olefins.

Hydrogen is used in the oxyhydrogen torch and in the atomic hydrogen torch to produce high temperatures for welding and cutting metals. It is used in the metallurgical industries to reduce metal oxides such as those of tungsten and molybdenum, to provide a reducing atmosphere in the heat-treatment of metals, and in the manufacture of metal hydrides. Although once used extensively to inflate dirigible balloons, it has been largely superseded by helium because of a number of disastrous explosions. When not used directly at the site of manufacture, hydrogen is transported and stored in steel cylinders at a pressure of 120–150 atmospheres (atm).

Natural occurrence. Hydrogen, in the free state, is only a minor component of the earth. It constitutes less than 1 part per million (ppm) of the atmosphere. It is found in the gases from some volcanoes, oil wells, and coal mines. It may be liberated as a product of the decomposition of organic matter and has been observed in the intestinal gases of animals. In the combined state, hydrogen makes up 0.76% of the weight of the earth's crust to make it the ninth most abundant element; 13.5% of the atoms of the earth's crust are hydrogen, exceeded in number only by oxygen and silicon. Most of this hydrogen is present in sea water, of which it constitutes 10.82% by weight. Other important occurrences are in minerals, as hydrates, in the hydrocarbons of petroleum deposits, and in the organic constituents of all living organisms.

Hydrogen is believed to constitute approximately 90% of the atoms in the universe. The thermonuclear energy produced by fusion reactions of hydrogen nuclei is the source of most of the energy radiated by the sun and other stars.

Physical properties. Ordinary hydrogen has an atomic weight of 1.0080, and a molecular weight of 2.0160. The gas has a density at 0°C and 1 atm of 0.08987 g/liter. Its specific gravity, compared to [illegible]

Hydrogen dissolves in water to the extent of 0.0214 volume per volume of water at 0°C, 0.018 vol at 20°C, and 0.016 vol at 50°C. It is somewhat more soluble in organic solvents, and 0.078 vol dissolves in 1 vol of ethanol at 25°C. Many metals adsorb hydrogen. Palladium is particularly notable in this respect, and dissolves about 1000 times its volume of the gas. The adsorption of hydrogen in steel may cause "hydrogen embrittlement," which sometimes leads to the failure of chemical processing equipment.

The hydrogen atom has an ionization potential of 13.54 volts. The hydrogen nucleus (proton, mass 1) has a spin of $\frac{1}{2}h$ and a magnetic moment of 2.79270 nuclear magnetons. Its absorption cross section for thermal neutrons is 0.332×10^{-24} cm².

The hydrogen molecule may exist in either of two forms, known as ortho- and parahydrogen. These distinct forms are possible because the nucleus of the hydrogen atom is spinning in a toplike manner and two atoms may combine with their nuclei spinning in the same direction (ortho) or in opposite directions (para). These spin isomers, as

Table 1. Properties of hydrogen

Melting point	−259.2°C
Boiling point at 1 atm	−252.8°C
Density of solid at −259.2°C	0.0866 g/cm³
Density of liquid at −252.8°C	0.0708 g/cm³
Critical temperature	−240.0°C
Critical pressure	13.0 atm
Critical density	0.0301 g/cm³
Specific heat at constant pressure	
Gas at 25°C	3.42 cal/(g)(°C)
Liquid at −256°C	1.93 cal/(g)(°C)
Solid at −259.8°C	0.63 cal/(g)(°C)
Heat of fusion at −259.2°C	14.0 cal/g
Heat of vaporization at −252.8°C	107 cal/g
Thermal conductivity at 25°C	0.000411 cal/(cm)(cm²)(sec)(°C)
Viscosity at 25°C	0.00892 centipoise

they are called, are ordinarily fairly stable but may be rapidly interconverted by a suitable catalyst, such as activated charcoal or platinized asbestos. The ratio of the two isomers in an equilibrium mixture varies markedly with the temperature: near the absolute zero, equilibrium hydrogen consists entirely of the para modification; at room temperature and above, parahydrogen constitutes 25% and orthohydrogen 75% of the mixture. Pure parahydrogen may be readily prepared by passing liquid hydrogen over activated charcoal. Pure orthohydrogen has been prepared only in small amounts by separating it from parahydrogen by processes such as thermal diffusion or gas chromatography. Physical properties of parahydrogen are measurably different from those of ordinary hydrogen. For example, parahydrogen melts at −259.34°C and boils at −252.90°C, whereas ordinary hydrogen melts at −259.21°C and boils at −252.77°C. The thermal conductivity of parahydrogen is markedly greater than that of the ortho form; the difference is used in analyzing mixtures of the two. When ordinary hydrogen is liquefied, the heat evolved during the slow conversion of the equilibrium mixture to parahydrogen is responsible for the evaporation of large amounts of liquid hydrogen during storage. These losses may be averted by catalytically converting all of the orthohydrogen to the para form during the liquefaction. The ortho-parahydrogen interconversion can be catalyzed by free hydrogen atoms. Measurement of the rate of interconversion permits the determination of hydrogen-atom concentrations in chemical reactions.

Chemical properties. At ordinary temperatures, hydrogen is a comparatively unreactive substance unless it has been activated in some manner, for example, by a suitable catalyst. At elevated temperatures it is highly reactive.

Although ordinarily diatomic, molecular hydrogen dissociates at high temperatures into free atoms according to the equation

$$H_2 \rightleftarrows 2H$$

The heat of dissociation at 25°C is 104.2 kcal/mole. The calculated percentage dissociation is 0.08 at 2000°K, 7.8 at 3000°K, 62.2 at 4000°K, and 95.5 at 5000°K. Atomic hydrogen is also produced when an electrical discharge is passed through hydrogen gas at low pressure, or when a mixture of hydrogen and mercury vapor is irradiated with light of wavelength 2537 A from a mercury arc.

Atomic hydrogen is a powerful reducing agent, even at room temperature. It reacts with the oxides and chlorides of many metals, including silver, copper, lead, bismuth, and mercury, to produce the free metals. It reduces some salts, such as nitrates, nitrites, and cyanides of sodium and potassium, to the metallic state. It reacts with a number of ele- [illegible]

With organic compounds, atomic hydrogen reacts to produce a complex mixture of products. With ethylene, C_2H_4, for example, the products include ethane, C_2H_6, and butane, C_4H_{10}. The heat liberated when hydrogen atoms recombine to form hydrogen molecules is used to obtain very high temperatures in atomic hydrogen welding.

Hydrogen reacts with oxygen to form water:

$$2H_2 + O_2 \rightarrow 2H_2O$$

The heat of reaction is 57.6 kcal/mole of hydrogen. At room temperature, this reaction is immeasurably slow, but is accelerated by catalysts such as platinum, or by an electric spark, and then may take place with explosive violence. In the absence of catalysts, reaction between hydrogen and oxygen becomes measurable at about 300°C and rapid above 500°C. The reaction is believed to take place in a series of steps constituting a chain reaction.

$$\begin{gathered} OH + H_2 \rightarrow H_2O + H \\ H + O_2 \rightarrow O + OH \\ O + H_2 \rightarrow OH + H \end{gathered}$$

The intense heat of the reaction is utilized in the oxyhydrogen torch for cutting and welding metals.

Hydrogen reacts less vigorously with the other group VI elements. The reaction with sulfur

$$H_2 + S \rightarrow H_2S$$

is exothermic by only 5 kcal/mole; the corresponding reactions with selenium

$$H_2 + Se \rightarrow H_2Se$$

and tellurium

$$H_2 + Te \rightarrow H_2Te$$

are endothermic.

The reactivity of hydrogen with the halogens decreases in the order fluorine, chlorine, bromine, iodine. The reaction with fluorine

$$H_2 + F_2 \rightarrow 2HF$$

is violent, even in the dark at −252°C. Properly controlled, it has been used to produce the extremely high temperature of 4000°C. The reaction with chlorine

$$H_2 + Cl_2 \rightarrow 2HCl$$

is slow under ordinary conditions, but may become explosive under the influence of light or heat. As in the case of oxygen, a chain reaction is involved:

$$\begin{gathered} Cl_2 + h\nu \rightarrow Cl + Cl \\ Cl + H_2 \rightarrow HCl + H \\ H + Cl_2 \rightarrow HCl + Cl \quad \text{etc.} \end{gathered}$$

With nitrogen, hydrogen undergoes the important reaction

$$N_2 + 3H_2 \rightarrow 2NH_3$$

to give ammonia. For the conditions required to bring about this reaction, see AMMONIA.

Hydrogen reacts at elevated temperatures with a number of metals, including lithium, sodium, potas-

sium, calcium, strontium, and barium, to give hydrides. For example,

$$H_2 + 2Li \rightleftharpoons 2LiH$$

See HYDRIDE, METAL.

The oxides of many metals are reduced by hydrogen at elevated temperatures either to the free metal or to lower oxides. Metals which can be produced from their oxides in this way include copper, silver, bismuth, mercury, tungsten, iron, nickel, and cobalt. Similarly, the chlorides of silver, copper, nickel, and cobalt react with hydrogen at 300–750°C to give hydrogen chloride and the free metal. Metallic bromides and iodides are more easily reduced than the chlorides.

Hydrogen reacts at room temperature with the salts of the less electropositive metals, such as gold, silver, copper, or mercury, in aqueous solution, and reduces them to the metallic state. These reductions are ordinarily quite slow, but are markedly accelerated if the hydrogen is adsorbed on platinum or palladium.

In the presence of a suitable catalyst, such as platinum, palladium, or nickel, hydrogen reacts with unsaturated organic compounds and adds to the double bond. With *ethylene, the reaction is*

$$H_2C{=}CH_2 + H_2 \rightarrow H_3C{-}CH_3$$

Aldehydes and *ketones* are similarly reduced to *alcohols*. Acetaldehyde, for example, is reduced to *ethyl alcohol*:

$$CH_3CH{=}O + H_2 \rightarrow CH_3CH_2OH$$

See HYDROGENATION.

Principal compounds. Hydrogen is a constituent of a very large number of compounds containing one or *more other* elements. Such compounds include water, *acids, bases*, most organic compounds, and many *minerals*. Compounds in which hydrogen is combined with a single other element are commonly referred to as *hydrides*. These may be divided into three *general classes*, the ionic or salt-like hydrides, the *covalent or molecular* hydrides, and the transitional metal hydrides.

With the halogens, *fluorine, chlorine*, bromine, and iodine, hydrogen forms hydrides containing [illegible] hanic acids, for example, hydrochloric acid. The hydrogen halides are gases at room temperature. Hydrofluoric acid (*HF*), is polymeric; the others are monomeric. Their *physical* properties are given in Table 2. The hydrogen *halides* dissolve in water to give strongly acid solutions. Hydrogen fluoride is commonly prepared from calcium fluoride and sulfuric acid:

$$CaF_2 + H_2SO_4 \rightarrow CaSO_4 + 2HF$$

Hydrogen chloride is similarly prepared from sodium chloride and sulfuric acid. Hydrogen bromide and hydrogen iodide may be prepared by direct union of the elements or by hydrolysis of the corresponding phosphorus trihalides, PBr_3 and PI_3.

Table 2. *Properties of hydrogen halides*

	Hydrogen fluoride, $(HF)_x$	Hydrogen chloride, HCl	Hydrogen bromide, HBr	Hydrogen iodide, HI
Melting point, °C	−83.1	−114.8	−86.9	−50.7
Boiling point, °C	19.5	−81.9	−66.8	−35.4
Density of liquid, g/ml	0.991	1.194	2.77	2.85
At temperature °C	19.5	−86	−67	−47

Table 3. Properties of hydrogen peroxide and water

	Water, H_2O	Hydrogen peroxide, H_2O_2
Melting point, °C	0	−0.9
Boiling point, °C	100	151.4
Density, g/ml	1.000 at 4°C	1.465 at 0°C

Hydrogen forms two compounds with oxygen, water (H_2O) and hydrogen peroxide (H_2O_2). The physical properties of these compounds are compared in Table 3. Hydrogen peroxide may be prepared by the hydrolysis of a metal peroxide

$$Na_2O_2 + 2H_2O \rightarrow H_2O_2 + 2NaOH$$

or by electrolysis reactions. It is unstable, and slowly evolves oxygen on standing. It is a powerful oxidizing agent, and reacts vigorously with many organic compounds.

With sulfur, hydrogen forms the compound hydrogen sulfide, H_2S, a colorless gas with the odor of rotten eggs. It may be prepared by the direct reaction of hydrogen with sulfur, or by treating a metal sulfide with an acid. A series of polysulfides of hydrogen, with the formulas H_2S_2, H_2S_3, H_2S_4, H_2S_5, and H_2S_6, are also known. With selenium and tellurium, hydrogen forms the compounds hydrogen selenide, H_2Se, and hydrogen telluride, H_2Te.

Boron forms a series of covalent hydrides, *the* best-known member of which is the gas, *diborane*, B_2H_6. Other members of the series are *pentaborane*, B_5H_9, a liquid, and decaborane, $B_{10}H_{14}$, *a crystalline* solid. Because of their high heats of *combustion*, these compounds and their derivatives *have received* considerable attention as possible *rocket fuels*.

The largest class of the covalent hydrides comprises the compounds of hydrogen with carbon, known as hydrocarbons. The first member of this series is methane, CH_4, a gas.

Among the hydrides of nitrogen are ammonia, NH_3, and hydrazine, N_2H_4. Ammonia is a colorless gas which melts at −78°C and boils at −33.3°C. Hydrazine is a liquid with a melting point of 1.4°C [illegible] H_4; phos- [illegible] H_3. [illegible] ds of hydrogen, *see* ACID AND BASE; HYDRAZINE; *HYDRO*CARBON; HYDROGENATION; PEROXIDE; SILICON; WATER.

Preparation. A large number of methods may be used to prepare hydrogen gas. The choice of method is determined by such factors as the quantity of hydrogen desired, the purity required, and the availability and cost of raw materials. Among the processes frequently used are the reactions of metals with water or acids, the electrolysis of water, the reaction of steam with hydrocarbons or other organic materials, and the thermal decomposition of hydrocarbons.

Small amounts of hydrogen are readily prepared in the laboratory by the reaction of zinc with dilute hydrochloric acid. The reaction is

$$Zn + 2HCl \rightarrow ZnCl_2 + H_2$$

The gas may be purified by passing it through an acidified solution of potassium permanganate or potassium dichromate and then through a solution of sodium hydroxide. It may be dried by passage through concentrated sulfuric acid or over silica gel. Other metals more electropositive than hydrogen will likewise liberate hydrogen gas upon treatment with water or acid solutions. Thus, metallic sodium reacts violently with water according to the equation

$$2Na + 2H_2O \rightarrow 2NaOH + H_2$$

The reaction may be moderated by amalgamating the sodium with mercury. The dissolution of iron filings in dilute sulfuric acid has also been frequently used as a source of hydrogen.

For the preparation of somewhat larger quantities of hydrogen, for example, for use in filling balloons, a convenient method is the reaction of calcium hydride with water:

$$CaH_2 + 2H_2O \rightarrow Ca(OH)_2 + 2H_2$$

In this reaction 1 kg of calcium hydride will produce about 1 m^3 of hydrogen. Another process sometimes used for the production of moderate amounts of hydrogen is the dissolution of aluminum or of silicon in alkali

$$2Al + 2NaOH + 6H_2O \rightarrow 2NaAl(OH)_4 + 3H_2$$
$$Si + 4NaOH \rightarrow Na_4SiO_4 + 2H_2$$

Hydrogen gas is produced on a large scale industrially. The amount produced annually in the United States is over 40,949,000,000 ft^3, excluding that consumed in the manufacture of ammonia and methanol. At present, the most important single process for hydrogen production is the steam-hydrocarbon process. The hydrocarbons may be either natural gas (largely methane) or refinery gases such as propane. The reactions involved, in the case of propane are as follows:

$$C_3H_8 + 3H_2O \rightarrow 3CO + 7H_2$$
$$3CO + 3H_2O \rightarrow 3CO_2 + 3H_2$$

The first reaction takes place when a mixture of propane vapor and steam is passed through a furnace at 1400–1800°F. After the temperature is lowered to 800°F and the gas mixture is passed through an iron oxide catalyst, the carbon monoxide reacts with additional steam to give carbon dioxide and more hydrogen. The second reaction is commonly referred to as the water-gas shift reaction.

Another important process for hydrogen production is the reaction of steam with water gas. Water gas, a mixture of carbon monoxide and hydrogen, is made by treating coke or coal with steam at a temperature of 1800°F or higher.

$$C + H_2O \rightarrow CO + H_2$$

Application of the water-gas shift reaction, as in the case of the steam-hydrocarbon process described above, yields hydrogen and carbon dioxide. The hydrogen obtained from these processes is purified by scrubbing with an aqueous solution of monoethanolamine.

Hydrogen is also produced industrially by the electrolysis of water containing dissolved potassium hydroxide. This process is comparatively expensive, but has the advantage of generating hydrogen of very high purity (over 99.9%). High-purity oxygen is a by-product. The over-all reaction is

$$2H_2O \rightarrow 2H_2 + O_2$$

Other important industrial processes are the reaction of steam with iron at 1500°F

$$Fe + H_2O \rightarrow FeO + H_2$$

the thermal dissociation of ammonia

$$2NH_3 \rightarrow N_2 + 3H_2$$

used for the production of comparatively small amounts of hydrogen for metal treating or for catalytic hydrogenation, and the catalytic reforming of petroleum, which yields hydrogen as a by-product. The increased use of the last-named process in recent years has resulted in the production of so much hydrogen that the market has become saturated and large amounts are burned as fuel.

Analytical methods. There are no fully satisfactory reagents for the direct absorption of hydrogen. Colloidal palladium absorbs large amounts of hydrogen, but the reagent is unstable, and it is necessary to remove carbon dioxide, olefins, and carbon monoxide before absorption of the hydrogen. Silver phosphate and silver borate have also been used to absorb hydrogen from a gas mixture.

The most commonly used methods for determining hydrogen in a gas mixture depend on burning the hydrogen to water and measuring the reduction in volume of the gas. The oxidation may be performed by adding excess oxygen and exploding the mixture either by passing an electrical spark between platinum electrodes immersed in the gas, or by passing the gas over a platinum wire heated to about 600°C. Since ½ mole of oxygen is required to burn 1 mole of hydrogen, the volume of hydrogen is equal to two-thirds the reduction in volume. Alternatively, the hydrogen may be oxidized by passing it through a tube filled with copper oxide heated to 290°C. The latter procedure is quite sp

cific for hydrogen, because the only other common gas oxidized by copper oxide at this temperature is carbon monoxide, the oxidation of which results in no volume change. Because no oxygen gas is consumed in this procedure, the volume of hydrogen is equal to the measured reduction in volume.

A number of physical measurements may be used to determine hydrogen, especially in mixtures with a single other gas whose identity is known. The exceptionally high thermal conductivity and low density of hydrogen make the measurement of either of these properties a sensitive indicator of hydrogen content. Other physical methods useful for hydrogen analysis include measurement of the velocity of sound in the gas and measurement of the refractive index. The hydrogen content of quite complex gas mixtures can often be determined by mass spectrometry or by gas chromatography.

The determination of chemically bound hydrogen, particularly in organic compounds, is usually performed by burning the compound in a stream of oxygen at about 700°C. The combustion water is absorbed in a dehydrating agent such as anhydrous magnesium perchlorate; the amount of hydrogen is calculated from the increase in weight of the absorption tube. [L.K.]

Bibliography: A. Farkas, *Orthohydrogen, Parahydrogen and Heavy Hydrogen*, 1935; D. T. Hurd, *An Introduction to the Chemistry of the Hydrides*, 1952; J. W. Mellor, *A Comprehensive Treatise on Inorganic and Theoretical Chemistry*, vol. 1, 1922; P. W. Mullen, *Modern Gas Analysis*, 1955; H. Remy, *Treatise on Inorganic Chemistry*, vol. 1, 1956; M. C. Sneed and R. C. Brasted, *Comprehensive Inorganic Chemistry*, vol. 6, 1957.

Hydrogen bomb

A device in which an uncontrolled, self-sustaining, thermonuclear fusion reaction is carried out in heavy hydrogen (deuterium) gas to produce an explosion. The hydrogen bomb is a fusion bomb. In a fusion reaction, the collision of two energy-rich nuclei results in a mutual rearrangement of their protons and neutrons to produce two or more reaction products, together with a release of energy of amount E given by Albert Einstein's formula $E = mc^2$, where m is the mass difference between the original and produced nuclei, and c is the velocity of light. *See* INERTIA OF ENERGY; *see also* FUSION, NUCLEAR; THERMONUCLEAR REACTION.

For the hydrogen bomb reaction to become self-sustaining, a so-called critical temperature of about 3.5×10^7°K must be attained, possibly with the aid of the enormous temperatures developed by the explosion of a fission bomb. Once this temperature is achieved, the energy released in the initial reactions maintains the temperature, and the chain proceeds either until the supply of fusionable material is exhausted, or until sufficient expansion has taken place that the gas is cooled below the critical temperature.

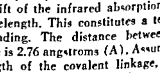
Yields of fusion bombs in excess of 10 megatons (4.18×10^{23} ergs) have been reported. *See* ATOMIC BOMB; NUCLEAR EXPLOSION. [K.W.P.]

Hydrogen bond

A bond formed by an atom of hydrogen which is initially attached to another atom by a strong covalent bond and which also is attracted to a third atom with sufficient force so that the three atoms are linked together in a chain. Since hydrogen can form only one covalent link, the attraction for the third atom must be essentially ionic in character. Only electronegative atoms such as fluorine, oxygen, and nitrogen form hydrogen bonds. For a hydrogen atom to bond readily, it must be of acidic character as is indicated by the electronegative character of the atoms joined. The atoms joined by a hydrogen bond usually belong to different molecules, but in cases where a six-membered ring can be formed, they may lie in a single molecule. When the atoms belong to the same molecule, the process is called chelation.

The third atom in the hydrogen bond, often termed the donor atom, usually has an unbonded pair of electrons. The so-called associated liquids owe their peculiar properties, such as high dielectric constant, to the formation of chains of molecules of the liquid by hydrogen bonding. The ionic character of the bond is strongly accentuated in the liquid of high dielectric constant so that the hydrogen may be thought of as a positive ion with a coordination number of 2. Examples are water, hydrogen fluoride, and hydrogen cyanide. In the case of the bifluoride ion (HF_2^-), there are three possible structures which may be in resonance, and the average position of the hydrogen may be thought of as intermediate between the two fluoride ions.

In the case of the carboxylic acids, a double hydrogen linkage leads to the formation of dimers with physical properties very different from the associated liquids mentioned above.



Dimer of acetic acid formed as a result of hydrogen bonding

The weakening of the covalent linkage when a hydrogen bond is formed causes a lowering of the stretching frequency of the hydrogen atom, resulting in a shift of the infrared absorption bands to longer wavelength. This constitutes a test for hydrogen bonding. The distance between oxygen atoms in ice is 2.76 angstroms (A). Assuming 1.0 A as the length of the covalent linkage, the ionic bond length would be 1.76 A. When a hydrogen bond is broken, the ionic portion breaks. The hydrogen-bond strength is only 5 or 6 kcal/mole as a rule.

Because of the small energy of activation, hydrogen bonds are readily formed and play an important role in fixing the structures of proteins in which chains are linked together either directly or through intermediate water molecules. Hydro-

gen bonds are involved in the action of adhesives such as mucilage and in the so-called hydration of paper. See CHELATION; CHEMICAL BINDING. [W.H.RO.]

Hydrogen electrode

A noble metal of large surface area covered with hydrogen gas in a solution of hydrogen ion saturated with hydrogen gas. The hydrogen gas may also be bubbled continuously over the noble metal; in this case, the affluent gas must be saturated with water vapor. Platinum is generally used as the noble metal. A large surface area is achieved by plating the surface with platinum sponge in a solution of chloroplatinic acid. In general, the noble metal is used in foil form and is welded to a wire sealed in the bottom of a hollow glass tube which is partially filled with mercury. Electrical contact to an external circuit is made through the mercury usually by a copper wire.

The hydrogen electrode is reversible to hydrogen ion and its electric potential is a logarithmic function of hydrogen-ion activity. Because of this relation, this electrode is used to measure hydrogen-ion activity. It cannot be used in reducing or oxidizing solutions or in solutions that poison the surface of the noble metal. Accordingly, it has limited application in the measurements of hydrogen-ion activity of solutions.

When the hydrogen-ion activity is unity and the hydrogen-gas pressure is at 1 atm, the potential of the hydrogen electrode is conventionally assigned a value of zero at all temperatures, and is used as a reference or standard electrode to which the potentials of all other electrodes are referred. All electrodes having a positive potential relative to the standard hydrogen electrode have greater reducing power than hydrogen; conversely, those of lower potential have a lower reducing power. Elements arranged in a series of their reducing power constitute what is called the electromotive-force series of the elements. See ELECTROCHEMICAL SERIES.

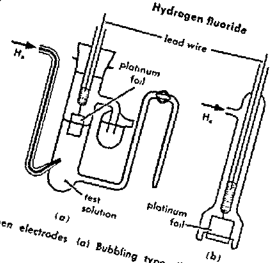
The potential of the hydrogen electrode varies with hydrogen-gas pressure p_{H} and with hydrogen-ion activity a_{H^+} according to the equation:

$$E = -\frac{RT}{F}\ln a_{H^+} + \frac{RT}{2F}\ln p_{H_2}$$

where R is the gas constant, T the absolute temperature, and F the faraday. At 25°C this relation is

$$E = -0.059158 \log a_{H^+} + 0.029579 \log p_{H_2}$$

Thus, for a 10-fold change in hydrogen-ion activity and a constant pressure of hydrogen gas, the potential of the hydrogen electrode changes by 0.059158 volt. At constant hydrogen-ion activity, a 10-fold change in hydrogen-gas pressure causes a change in the hydrogen electrode potential of 0.029579 volt. In determining the pressure of the hydrogen gas, the vapor pressure of the solution p should be subtracted from the observed barometric pressure P. The observed potential of the electrode is corrected to 1 atm hydrogen pressure by the relation:



Hydrogen electrodes (a) Bubbling type. (b) Dipping type.

$$E_{1\,atm} = E_{obs} + \frac{RT}{2F}\ln\frac{760}{(P-p)}$$

For approximate measurements at room temperature, corrections for the vapor pressure of the solution may be neglected.

The hydrogen electrode may be used at atmospheric pressure from 0 to 100°C and at higher temperatures for pressures exceeding 1 atm. It may also be used in nonaqueous solutions which do not have reducing or oxidizing power on the electrode.

The electrode may be prepared in three general forms: (1) dipping type, (2) stationary bubbling type, and (3) rocking type. In the dipping type, the gas is introduced by a tube adjacent to the hydrogen electrode; both are supported at the top by a stopper or other means. The dipping type may also be of the type shown in the illustration. In the stationary bubbling type, the gas is introduced below the hydrogen electrode (see illustration), and the vessel is provided with a side arm whereby it may be joined to another electrode or half-cell. In the rocking type, the solution is agitated and made to move over the hydrogen electrode in a rocking motion; contact to another electrode or half-cell is made through a stopcock and a flexible rubber connection. This type is sealed, and therefore, has limited life. See ELECTRODE POTENTIAL; HYDROGEN ION. [W.J.H.]

Hydrogen fluoride

The hydride of fluorine and the first member of the family of halogen acids. Anhydrous hydrogen fluoride is a mobile, colorless liquid that fumes strongly in air. It has the empirical formula HF, melts at −83°C, and boils at 19.8°C. Hydrogen fluoride is prepared on the large industrial scale by treating fluorspar (calcium fluoride, CaF_2) with concentrated sulfuric acid. The crude product is purified by fractional distillation to yield a product containing more than 99.5% HF; the remaining impurities are principally water and small amounts of sulfur dioxide, silicon tetrafluoride, and boron trifluoride. Very dry hydrogen fluoride can be obtained by electrolysis or by treatment with r

agents such as fluorine or cobaltic fluoride that react with water.

Properties. Anhydrous hydrogen fluoride is an extremely powerful acid, exceeded in this respect only by 100% sulfuric acid. Like water, hydrogen fluoride is a liquid of high dielectric constant that undergoes self-ionization and forms conducting solutions with many solutes. Alkali metal fluorides and silver fluoride dissolve readily in hydrogen fluoride to form conducting solutions. The alkali metal fluorides are bases in the hydrogen fluoride system and correspond to solutions of alkali metal hydroxides in water. Conversely, antimony pentafluoride and boron trifluoride act as acids in hydrogen fluoride and accentuate the already strong acid properties of the solvent.

Anhydrous hydrogen fluoride dissolves a wide variety of organic compounds, actually many more than are soluble in water. Oxygen-, nitrogen-, and sulfur-containing compounds usually have high solubility in liquid hydrogen fluoride. Aromatic hydrocarbons are moderately soluble, and even saturated aliphatic hydrocarbons show appreciable solubility. Despite the fact that hydrogen fluoride is a strong dehydrating agent, many organic solutes can be recovered unchanged from hydrogen fluoride solution. Even sensitive compounds, such as carbohydrates and proteins, which are highly soluble in hydrogen fluoride, can under appropriate conditions be recovered from solution quite unchanged.

Uses. Hydrogen fluoride is a widely used industrial chemical. It was formerly used in the petroleum refining industry for the isomerization of aliphatic hydrocarbons to form more desirable automotive fuels, but this application has been superseded by other methods. The largest industrial use of hydrogen fluoride is in the preparation of fluorine-containing refrigerants (freons, genetrons). An increasingly important use of hydrogen fluoride is in the preparation of organic fluorocarbon compounds by the Simons electrochemical process. In this procedure, an organic compound is dissolved in hydrogen fluoride, and an electric current passed through the solution, whereupon the hydrogen atoms in the organic [illegible] fluoride is used for the conversion of uranium dioxide to uranium tetrafluoride, an intermediate in the preparation of uranium metal and uranium hexafluoride. Important organic reactions may in some cases be performed to advantage in hydrogen fluoride solution, and nitration, sulfonation, diazotization, cyclization, and polymerization reactions have been carried out in this medium. The important metalloorganic compound *ferrocene* (iron dicyclopentadienyl) can be conveniently prepared in hydrogen fluoride.

Aqueous solutions of hydrogen fluoride (hydrofluoric acid) are relatively weakly acidic as compared to hydrochloric acid. Fluoride salts are formed by reaction of hydrofluoric acid with metal oxides and carbonates. Of particular importance is the rapid reaction of hydrofluoric acid or anhydrous hydrogen fluoride with silica, which leads to the application of these substances as etching agents for glass.

Both hydrogen fluoride and hydrofluoric acid cause unusually severe burns; appropriate precautions must be taken to prevent any contact of the skin or eyes with either the liquid or the vapor. *See* FLUORIDE; FLUORINE; HALOGENATED HYDROCARBON; ISOMERIZATION. [J.J.K.]

Bibliography: J. H. Simons (ed.), *Fluorine Chemistry*, vol. 1, 1950.

Hydrogen ion

A proton combined with a molecule of water, formula H_3O^+. It is often called the hydronium or oxonium ion, and it is found in pure water and in all aqueous solutions. For simplicity, the hydrogen ion is usually designated by the symbol H^+. The hydrogen ion is always associated with the hydroxyl ion in aqueous solutions, and the products of the concentrations, or more precisely, the activities, of these two ions are equal to a constant known as the ion-product constant of water. Its numerical value increases rapidly with increasing temperature and, at 25°C, is given by the expression

$$K_{H_2O} = [H^+][OH^-] = 1 \times 10^{-14} \qquad (1)$$

wherein $[H^+]$ and $[OH^-]$ are the concentrations of the hydrogen and hydroxyl ions in moles per liter.

In pure water, the ions are produced by the interaction of two molecules of water, one donating a proton to the other in accordance with the equation,

$$H_2O + H_2O \rightleftharpoons H_3O^+ + OH^-$$

According to Eq. (1), in a neutral water solution, each ion would have a concentration of 1×10^{-7} mole/liter. Whether a solution is acidic or basic depends upon the relative concentrations of the two ions. In an acid solution, the hydrogen-ion concentration is greater than 1×10^{-7} mole/liter. In a basic solution, it is less, with a corresponding increase in hydroxyl-ion concentration. Hydrogen ions are formed when any proton donor reacts with water. For example, with an acid or an acid ion the following reactions take place:

$$HCl + H_2O \rightleftharpoons H_3O^+ + Cl^-$$
$$H_2PO_4^- + H_2O \rightleftharpoons HPO_4^{--} + H_3O^+$$

Hydrogen ions are also formed in cation hydrolysis.

The hydrogen ion takes part in many reactions. Among the more important of these are the neutralization of hydroxide bases to form water and the interaction with bases such as ammonia to form the corresponding cation. Interaction with carbonates and bicarbonates, sulfites and bisulfites, and similar salts produces the volatile gases carbon dioxide and sulfur dioxide, respectively. The hydrogen ion will react with metals above hydrogen in the activity series to form hydrogen gas and the ion of the metal.

Hydrogen ions determine the course of many reactions occurring in living organisms and in the chemical processes of industry. The close control of hydrogen-ion concentration through the use of buffer systems is of primary importance in life processes. The control of acidity is of importance in sanitary engineering and electroplating, and in the textile, pharmaceutical, and food industries. Hydrogen-ion concentration plays a major role in determining the potential of most oxidation-reduction systems. Solutions can be made highly conducting by the addition of acids because of the high conducting power of the hydrogen ion.

pH. The hydrogen-ion concentration and the hydrogen-ion activity of most solutions is a small number that can be expressed in exponential form. In order to eliminate the use of exponential numbers in expressing hydrogen-ion concentration, S. P. L. Sorensen, in 1909, introduced the concept of pH which is defined as follows:

$$\mathrm{pH} = \log \frac{1}{[\mathrm{H}^+]} \tag{2}$$

wherein $[\mathrm{H}^+]$ is the actual concentration of the hydrogen ion in molarity units. Since all aqueous solutions contain both hydrogen and hydroxyl ions, and since the product of their concentrations is equal to a constant, it is possible to express all degrees of acidity and basicity on the pH scale. The table gives a series of values for the hydroxyl-ion concentrations and the hydrogen-ion concentrations, with the corresponding pH values, for a series of solutions at 25°C.

Relation between ion concentrations and pH at 25°C

$[\mathrm{H}^+]$	$[\mathrm{OH}^-]$	pH	$[\mathrm{H}^+]$	$[\mathrm{OH}^-]$	pH
1	1×10^{-14}	0	1×10^{-7}	1×10^{-7}	7
1×10^{-1}	1×10^{-13}	1	1×10^{-8}	1×10^{-6}	8
1×10^{-2}	1×10^{-12}	2	1×10^{-10}	1×10^{-4}	10
1×10^{-4}	1×10^{-10}	4	1×10^{-12}	1×10^{-2}	12
1×10^{-6}	1×10^{-8}	6	1×10^{-14}	1	14

This table brings out certain characteristics of the pH unit. A neutral solution has a pH of 7, acid solutions have a pH of less than 7, and basic solutions have a pH greater than 7. A change of one pH unit in acidity corresponds to a 10-fold change in the hydrogen-ion concentration. Solutions seldom have a hydrogen-ion concentration greater than 1 mole/liter, but such a solution would have a negative value for the pH. The application of Eq. (2) shows that the pH of a solution whose hydrogen-ion concentration is 2.00×10^{-3} *M* would have a pH of 2.70, and a solution whose hydroxyl-ion concentration is 2.00×10^{-3} *M* has a pH of 11.30. In practice, the pH of a solution is seldom expressed to more than two decimal places.

In the more precise treatment of solutions, activities are used in place of concentrations, and the term paH is often used. This is equal to the logarithm of the reciprocal of the hydrogen-ion activity. Most of the present methods for determining hydrogen-ion concentration really determine the activity. However, if only three significant figures are used, pH and paH can be used interchangeably without serious error.

The hydrogen ion is the most highly conducting of all ions. With a potential gradient of 1 volt/cm at 25°C, the hydrogen ion is assigned a speed of 0.00362 cm/sec. This speed is almost five times that of the next most highly conducting cation. Other ions move through the solution as individual entities, whereas with the hydrogen ion, in many but not all solvents, the proton is passed from one solvent molecule to another in a chainlike manner. This explains the high conductance of acid solutions containing high concentrations of hydrogen ions.

Thermodynamic values. In the compilation of thermodynamic data tables, a value of zero is assigned to the standard free energy, standard enthalpy, and standard entropy of the hydrogen ion. This value applies to solutions in which the activity of the hydrogen ion is unity. The values of these three thermodynamic functions for other ions are then expressed in terms relative to the hydrogen ion. In like manner, the potential of the standard hydrogen-gas electrode is arbitrarily assigned a value of zero, and all other electrode potentials are expressed in terms relative to the hydrogen electrode. The standard hydrogen electrode is one in which the hydrogen gas has a pressure of 1 atm and the hydrogen ions are at unit activity. Various thermodynamic equilibrium constants such as the ionization constants of acids and bases, hydrolysis constants, and indicator constants may be determined through the use of hydrogen-ion activities.

Determination of hydrogen-ion activities. The two general methods available for the determination of hydrogen-ion activities and the pH values of solutions are the potentiometric procedures and the colorimetric procedures. Titration does not give the actual concentration or activity, but rather the total available acidity of a solution. In the potentiometric procedures, the more important pH-indicating electrodes are the hydrogen-gas electrode, the quinhydrone electrode, and the glass electrode. Each of these, in combination with a reference electrode, constitutes a chemical cell whose potential can be used to determine [illegible] glass electrode, in combination with reference electrodes, is the cell most used for pH measurements. Compact units with the accompanying electrical measuring circuits are termed pH meters and are available commercially.

Colorimetric procedures are based on the fact that definite hydrogen-ion concentrations give definite shades of color to solutions of various organic substances termed acid-base indicators. These indicators exist in two color forms, and the color change in these substances is gradual over a pH range of about two units. Consequently, for each indicator, it is possible to prepare color standards

which can be matched against solutions of unknown pH in order to determine the acidity of these solutions. Color-matching procedures can be carried out in an approximate manner using a tube with an unknown pH solution and a series of tubes containing solutions of known pH and the same concentration of indicator. *See* ACID AND BASE; BUFFER SOLUTION; ELECTROCHEMICAL SERIES; ELECTROLYTIC CONDUCTANCE; EQUILIBRIUM, IONIC; INDICATOR, ACID-BASE; OXIDATION-REDUCTION; VOLUMETRIC ANALYSIS. [H.D.C.; S.B.K.]

Bibliography: H. T. S. Britton, *Hydrogen Ions*, 2 vols., 4th ed., 1956.

Hydrogenation

The chemical reaction of substances with molecular hydrogen in the presence of a catalyst. The process includes reactions involving cleavage of molecules (hydrogenolysis), syntheses in which hydrogen simply adds to reactive molecules, and also reactions such as isomerization and cyclization which occur in the presence of hydrogen and a catalyst. Other reactions that involve molecular hydrogen and catalysts are reductive amination (hydroammonolysis) and hydroformylation (Oxo reaction)

In the exact usage of the term, hydrogenation is synonymous with reduction, which applies to those reactions wherein oxygen or some other element (most commonly nitrogen, sulfur, carbon, or halogen) is withdrawn from, or hydrogen is added to, a molecule. When catalytic hydrogenation is capable of producing the desired reduction product, it is generally the simplest and most efficient procedure.

Catalytic hydrogenation is used extensively in industrial processes. Important examples are the synthesis of methanol, liquid fuels, hydrogenated vegetable oils, fatty alcohols from the corresponding carboxylic acids, alcohols from aldehydes prepared by the aldol reaction, cyclohexanol and cyclohexane from phenol and benzene, respectively, and hexamethylenediamine for the synthesis of nylon from adiponitrile.

Hydrogenolysis. This term refers particularly to cleavages in a molecule associated with the addition of hydrogen. Hydrogenolysis is analogous to hydrolysis and ammonolysis, which involve cleavage of a bond induced by the action of water and ammonia, respectively. Examples of bonds that are broken by hydrogenolysis reactions are: carbon-carbon, carbon-oxygen, carbon-sulfur, and carbon-nitrogen.

Catalytic hydrogenation. Acetylenes readily add two moles of hydrogen under catalytic conditions to give the saturated derivatives. With proper conditions, the hydrogenation can be stopped at the intermediate olefin stage:

$$RC{\equiv}CR' + H_2 \rightarrow RCH{=}CHR' + H_2 \rightarrow RCH_2CH_2R'$$

where R and R′ are aliphatic, aromatic, or certain other groups.

Olefins readily undergo hydrogenation, usually in the presence of a nickel-containing catalyst and sometimes with platinum or palladium catalysts. Many of the lower-molecular-weight olefins such as ethylene, propylene, butenes, pentenes, and hexenes are readily hydrogenated in the vapor phase by passing the olefin and hydrogen over a nickel-containing catalyst at atmospheric pressure and 100–200°C. Higher-molecular-weight olefins are generally hydrogenated in the liquid phase.

Aromatic compounds may be reduced either in the vapor phase at atmospheric pressure or in the liquid phase at hydrogen pressures of 100–200 atm. In the latter case aromatics, such as benzene, toluene, and *p*-cymene, can be hydrogenated readily in the presence of a nickel catalyst. In the case of naphthalene or substituted naphthalenes, the product may be the tetra- or decahydronaphthalene derivative.

Carbonyl compounds, such as aldehydes and ketones, are easily reduced by catalytic hydrogenation to the corresponding alcohols according to the equation

$$RCOR' + H_2 \rightarrow RCHOHR'$$

where R is an aliphatic or aromatic group, and R′ may be the same group or a hydrogen atom. Frequently when R is an aromatic group it is difficult to stop the reduction at the alcohol. Instead, it proceeds further to yield a hydrocarbon, RCH_2R'. In general, aldehydes are reduced more rapidly than ketones, although there are numerous examples in [illegible] temperature [illegible] essure [illegible] metal [illegible] catalysts.

Esters and acids react with hydrogen in the presence of catalysts according to the general equation

$$RCOOR' + 2H_2 \rightleftharpoons RCH_2OH + R'OH$$

where R is usually an aliphatic group and R′ is an aliphatic group or a hydrogen atom. The reaction is reversible, with the temperature and pressure determining the relative equilibrium proportions.

Mechanism of hydrogenation. Hydrogenation reactions are generally reversible. Catalysts affect the rate or speed of reaction, but have nothing to do with the inherent tendency of the reaction to proceed. To know whether or not the reaction is at all possible, it is necessary to determine the free energy change of a definite, specific reaction. For example, the thermodynamic equations pertaining to the change in free energy content per atom of carbon in each molecule during the reduction of carbon monoxide at 400°C are:

$$\begin{aligned} CO(g) + 3H_2(g) &\rightleftharpoons CH_4(g) + H_2O(g) \\ -40{,}500 + 0 + \Delta F &= -3{,}600 - 50{,}000 \\ \Delta F &= -13{,}100 \text{ cal at } 400°C \end{aligned}$$

Here there is a free energy decrease which shows that, at this temperature, reduction of carbon monoxide to methane is possible. At 1000°C, the re-

verse reaction for the production of hydrogen from steam and natural gas is practiced on a large scale. The relationships for the hydrogenation of carboxylic acids to alcohols are

$$CH_3COOH(g) + 2H_2(g) \rightleftharpoons C_2H_5OH(g) + H_2O(g)$$
$$2(-37{,}000) + 0 + \Delta F = 2(-8{,}100) - 50{,}000$$
$$\Delta F = +7{,}200 \text{ cal at } 400°C$$

Although there is an actual, small free energy increase, this reaction proceeds with a decrease in volume so that equilibrium yields can be definitely helped by carrying out such a hydrogenation at relatively high pressures. *See* FREE ENERGY.

Effect of temperature. Practically every hydrogenation can be reversed by increasing the temperature. As a practical measure, it is essential to operate at as low a temperature as possible compatible with a satisfactory rate of reaction. In general, the temperatures for hydrogenation reactions are below 400°C.

Effect of pressure. Hydrogenation rates are generally increased by increasing the hydrogen pressure. Pressure also increases the equilibrium yield in hydrogenations where there is a decrease in volume as the reaction proceeds.

Catalysts. For the most part, hydrogenation catalysts are solids consisting of metals and metal oxides. Hydrogenation catalysts may be classified in accordance with their customary use. Vigorous catalysts suitable for the hydrogenation of alkyne and alkene linkages, aldehydes, and ketones include nickel and cobalt, and molybdenum and tungsten oxides or sulfides. Mild catalysts useful for stepwise hydrogenations of aldehydes and ketones include oxides of copper, zinc, and chromium, and metallic platinum and palladium. Molybdenum sulfide and especially tungsten disulfide are active catalysts for operations at 3000 psi. These catalysts are useful for the hydrogenation of unsaturates and to effect the cleavage of C—C, C—O, and C—N bonds.

Equipment. There are two types of reaction vessels. The first is used with liquids or solids, as in the hydrogenation of viscous hydrocarbons or in fat hardening, where internal agitators bring about an intimate mixture of organic compound, catalyst, and hydrogen. Usually, these are batch operations. The second type is used where the organic compound has sufficient vapor pressure at the reaction temperature, as in the synthesis of methanol from carbon monoxide, to permit gas-phase, continuous operations.

Gas-phase operations under extreme conditions of temperature and pressure require thick-walled pressure vessels fabricated from alloys that are impervious to, and unaffected by, hydrogen gas. Chrome-vanadium steel and BTG metal (12% chromium, 60% nickel, 2.5% tungsten) have been found satisfactory.

Typical hydrogenation processes. Hardening of animal fats and vegetable oils is carried out on a large scale to produce products of desired consistency and to remove certain impurities. Chemically, the process involves the conversion of glycerides of unsaturated fatty acids (for example, oleic and linoleic) to saturated ones in the presence of a nickel catalyst. The conversion of olein to stearin may be expressed as follows:

$$(C_{17}H_{33}COO)_3C_3H_5 + 3H_2 \rightarrow (C_{17}H_{35}COO)_3C_3H_5$$

The synthesis of methanol from carbon monoxide is a reversible reaction. The temperature range over which the reaction is practical is small. Below 300°C, the rate is slow; above 400°C, the equilibrium becomes unfavorable. Satisfactory conversions can be obtained at high pressures (3500 psi) because the reaction involves a decrease in volume.

$$CO + 2H_2 \rightleftharpoons CH_3OH$$

Metal and metal oxide mixtures, such as copper with oxides of zinc, chromium, or manganese, are useful catalysts.

Unlike the hardening of fats, which involves only the hydrogenation of ethylene linkages, the hydrogenolysis of the carboxyl group of acids and esters takes place with the formation of alcohols.

$$RCOOH + 2H_2 \rightarrow RCH_2OH + H_2O$$

The olefinic bonds in the fatty chain may or may not be reduced. A reduced copper-ammonium chromate catalyst is used at 350–400°C.

Petroleum, tar, and coal are hydrogenated to (1) improve existing products, (2) convert low-grade materials such as heavy oils into valuable fuels, and (3) transform solid fuels such as lignites and coal into liquid fuels. By a proper selection of catalysts and operating conditions, such hydrogenations can be directed to give desired end products and at the same time cause impurities that are common catalytic poisons, such as sulfur, nitrogen, and oxygen, to be detached from their molecular linkages and to be removed as hydrogen sulfide, ammonia, and water. *See* ADDITION REACTION; AMINATION; DEHYDROGENATION; FISCHER-TROPSCH PROCESS; HIGH-PRESSURE PROCESSES; HYDROFORMYLATION; HYDROGEN; ORGANIC CHEMICAL SYNTHESIS; OXIDATION-REDUCTION. [P.H.G.]

Bibliography: P. H. Groggins, *Unit Processes in Organic Synthesis*, 5th ed., 1958.

Hydrography

A division of science dealing with certain aspects of all the waters of the earth's surface, particularly charting them and describing their physical features and conditions: the position of lakes and oceans; the contours of the sea bottom; the position of shallows, deeps, and reefs; and the direction and volume of currents.

The term hydrography is now used less than formerly because the relatively narrow objectives of this science are becoming recognized as parts of the broader fields of hydrology and oceanography. However, the charting and navigational aspects continue strongly developed.

Charting the ocean bottom and shores, as well as lakes and navigable rivers, is of utmost importance to coastal and inland navigation and to the fishing industry. Every maritime nation of any importance has established as one of its regular bureaus a hydrographic office. The Hydrographic Department of the British Admiralty was established in 1795; the Hydrographic Office of the U.S. Navy was established in 1866. The United States also maintains the Coast and Geodetic Survey, whose duties include mapping the shorelines of the United States; and the Lake Survey of the U.S. Army Corps of Engineers, whose duties include mapping the Great Lakes and their interconnecting waterways. *See* OCEANS AND SEAS; OCEANOGRAPHY. [A.N.S.]

Hydroida

An order of the coelenterates which includes the fresh-water hydras, the attached and usually colonial hydroids, and many of the smaller jellyfish. It is the largest order of the class Hydrozoa.

Taxonomy. The order Hydroida is divided into two suborders, each with two names: one used for the hydroids and the other for the medusae. The Gymnoblastea includes hydroids without protective cups around the hydranths and gonozooids. Jellyfish of these athecate hydroids are Anthomedusae.

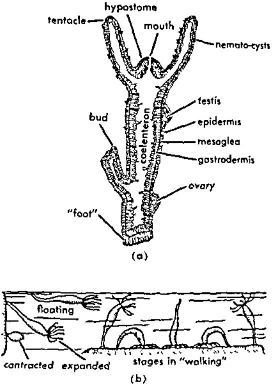
The suborder Calyptoblastea includes the hydroids with protective cups around the hydranths (*hydrothecae*) and around the gonozooids (gonothecae). Jellyfish of thecate hydroids are called Leptomedusae.

Morphology. Anthomedusae are typically ovoid jellyfish, often with eyespots. Leptomedusae are usually flattened or saucer-shaped, have statocysts (sense organs of balance), and lack eyespots. The gonads of Anthomedusae are generally on the wall of the stomach just above the mouth, and those of Leptomedusae below the radial canals.

Young hydranths of gymnoblastic hydroids are small with five or more tentacles, but subsequently grow much larger and add more tentacles. Calyptoblastic hydranths, in contrast, emerge from a bud with a full complement of parts; they do not grow, live for only about a week, undergo regression and absorption by the colony, and are then replaced by new hydranths.

The fresh-water hydras are simple, motile polyps which do not produce colonies. Buds separate from the parent and become individual polyps. Simple gonads develop on the body and there is no medusa. Hydras are sometimes included in the Gymnoblastea and sometimes placed in a suborder by themselves, the Hydrida.

Life cycle. The fertilized egg usually develops into a free-swimming ciliated larva, the planula, which soon attaches to some support and develops a mouth surrounded by tentacles at its free end. This attached stage is called a polyp and produces stolons, stems, and further polyps which remain connected to make up a hydroid or hydroid colony. Such colonies may show elaborate patterns of branching and may consist of hundreds of polyps, also called hydranths. Buds develop on the colony and are liberated as jellyfish or medusae which produce the sperm and eggs. *See* INVERTEBRATE EMBRYOLOGY.



Hydra. (a) Longitudinal section of *Hydra*. (b) Movements of *Hydra*. (From *T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957*)

Although the life cycle described for *Obelia geniculata* is regarded as typical, many species show reduction of either the polyp or the medusa stage. A comparison showing the differences in the life cycles of *Hydra* and *Obelia* follows. *Hydra*, a solitary polyp, has no medusoid stage, while *Obelia*, a polymorphic colony of small polyps, has a medusoid stage.

Hydra:

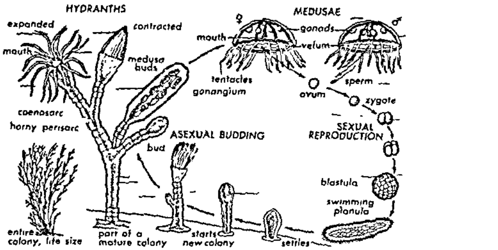
Polyp (1) → bud (2) → polyp (3) → egg / sperm (4) → zygote (5) → polyp (1)

Obelia:

Hydranth (1, feeding polyp) → bud (2) → hydranth (3) → gonangium (4, reproductive polyp) →

medusae (5, male or female) → egg / sperm (6) → zygote (7) → blastula (8) →

planula larva (9) → polyp (10) → polyp colony with hydranths (1, by budding)



Life cycle of Obelia. (From T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957)

The colony. In a hydroid colony there is sometimes only one kind of polyp and the medusae are developed on the hydranth or stem. Often, however, there are special reproductive polyps, with partial or complete reduction of mouth and tentacles. Such reproductive polyps are called *gonozooids* or *gonangia* and may produce small medusae which are liberated to grow and reproduce sexually. Often the medusa is incompletely developed except for its gonads, and the gametes ripen while the medusoid body remains attached to the parent hydroid.

In addition to ordinary nutritive hydranths and reproductive ones, some hydroids have other specialized types which are sensory, defensive, or aid in capturing food.

The stems of a colony are covered by a stiff secreted substance, the perisarc or periderm, which supports and protects it. The periderm may form cups around the hydranths and gonozooids.

Ecology. Hydroids are species in which the polyp stage is usually dominant. Most hydroids are found near the shore attached to various supports such as rocks, wharves, boats, mussels, barnacles, worm tubes, crab and snail shells and seaweeds. Few occur on mud or sand. Sometimes the medusa stage is well developed and the hydroid stage may be lacking. Medusae are abundant both in coastal waters and in the open sea.

Some hydroids live in direct association with other animals. A few are parasitic on fish, others live inside clams, on snails, hermit crabs and worm tubes. The larger animal is valuable to the hydroid because its activity produces a movement of water past the hydroid and thus suspended food is made available.

quickly replaced, a bit of stem can produce a new hydranth, and completely disorganized masses of cells can reconstitute a new polyp. See REGENERATION (BIOLOGY).

Dried Sertularia, a large and delicately branched hydroid, is sold as decorative material known as white weed (air fern when dyed green). See HYDROZOA. [S.CR.]

Bibliography: C. M. Fraser, *Hydroids of the Pacific Coast of Canada and the United States*, 1937; C. M. Fraser, *Hydroids of the Atlantic Coast of North America*, 1944; L. H. Hyman, *The Invertebrates*, Vol. 1, 1940; F. S. Russell, *The Medusae of the British Isles*, 1953.

Hydrokinematics

Motion of a liquid apart from the cause of the motion. A liquid is treated as a continuum, the elements of which move along continuous paths. At any instant the flow pattern may be delineated by a family of streamlines which are everywhere tangent to the paths, but do not in general coincide with them unless the flow is steady (see STOKES STREAM FUNCTION). Complete specification of a flow field requires that the distributions of sources and vorticity be known. See FLOW OF FLUIDS; HYDRODYNAMICS. [L.L.A.]

Hydrokinetics

Forces produced by a liquid as a consequence of its motion. The mass of a liquid gives rise to inertia forces which are manifested as normal stresses or pressures when the liquid flows. The viscosity and cohesion of a liquid give rise to tangential shear forces and also affect the normal stresses (see FLUID-FLOW PROPERTIES). The interaction of these properties on the motion of the fluid is formally expressed in the equations of Navier and Stokes. Because of the difficulty of solving these equations in general, approximate solutions are often ob-

tained by neglecting the less important properties in a particular flow situation. *See* HYDRODYNAMICS. [L.LA.]

Hydrolaccolith

A physiographic feature found in permafrost regions. The term is commonly applied to what are called oversize frost heaves (a few tens of feet and larger in diameter) or to a frost mound (upwarp of ground) produced by the freezing of water along a subsurface discontinuity which results in a lenticular body of ice. In a general way, the ice structure resembles an igneous laccolith. An alternate term, pingo (from an Eskimo name for conical hill), is sometimes used as a local name for frost mounds formed in this manner. The term pingo, however, is commonly restricted to frost mounds that are of longer than seasonal duration and that are, as a rule, of relatively large dimension, such as hundreds of feet and even a few thousands of feet in diameter *See* PERMAFROST. [M.M.M.]

Bibliography: E. deK. Leffingwell, *The Canning River Region, Northern Alaska*, USGS Profess. Paper 109, 1919; S. W. Muller, *Permafrost*, USGS and U.S. Army Corps of Engineers, Strategic engineering study, Spec Rept. 62, 1945 (reprint 1947).

Hydrology

The science of water, its properties, distribution, and circulation in the lands of the earth. The term refers to water on the surface of the land, in the soil, and in underlying rocks and to related aspects of water in the atmosphere which deal with evaporation and precipitation at the face of the earth.

Water is essential in all life processes including growth, for domestic, agricultural, and industrial uses, and for many recreational activities. Thus the study of water, its distribution and conservation, and the means by which it may be obtained and controlled for use is of utmost importance to the welfare of mankind.

On the other hand, water is often a destructive agent of awesome force. Great floods periodically inundate valleys, and storms lash coasts, causing death and destruction to man and his works. Less dramatic but nevertheless costly are the rising water tables which, especially in irrigated areas, cause deterioration of the soil and make worthless large areas that would otherwise produce bountiful crops, and the subsidence of the land surface in some areas of heavy withdrawal from wells. Hydrology has contributed substantially to the mitigation of the effects of water's destructive powers.

Modern hydrology gains stature among sciences by a shift in emphasis from qualitative to quantitative studies. As a result, the International Association of Scientific Hydrology is now a section of the International Union of Geodesy and Geophysics, and a Section of Hydrology has been established within the American Geophysical Union since 1930 (*see* GEOPHYSICS). Like other earth sciences, hydrology uses as its tools the basic sciences of mathematics, physics, and chemistry. However, the lines of distinction between it and the sciences of geology, meteorology, oceanography, and geochemistry are seldom sharp, but each has a particular focus of investigation.

The central concept of hydrology is the hydrologic cycle, a term used to describe the circulation

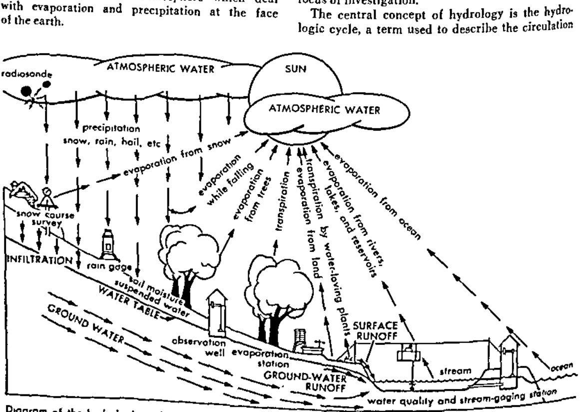


Diagram of the hydrologic cycle.

of water from the oceans through the atmosphere to the land and back to the oceans over and under the land surface. Water travels over sea and land as part of the vast air masses moving around the earth. It condenses into clouds and is precipitated as rain, snow, or sleet when one air mass rises to pass over another or over a mountain. Some of that which falls is evaporated while still in the air and some falls on vegetation to be evaporated directly from it, and so complete the cycle very quickly. Of the precipitation that reaches the ground surface, some evaporates quickly, some penetrates the soil, and some runs off over the land surface into streams, lakes, or ponds. Of that which penetrates the soil, some is held for a time and then returned to the atmosphere by evaporation or plant transpiration. The remainder penetrates below the soil zone to become a part of the ground water. Study of the water in the atmosphere, although closely related to hydrology, lies more properly in the field of meteorology (*see* METEOROLOGY). The science relating to the oceans also touches closely on hydrology but is regarded as a separate science (*see* OCEANOGRAPHY).

Hydrology is especially concerned with water after it is precipitated on the continents and before it returns to the oceans. It is concerned with amounts and intensity of precipitation; quantities of water stored as snow and in glaciers, and rates of advance or retreat of glaciers; discharge of streams at various points along their courses; gains and losses of water stored in lakes and ponds; rates and quantities of infiltration into the soil and movement of soil moisture; changes of water levels in wells as an index of gains and losses in storage of ground water; occurrence and rate of movement of water in underground reservoirs; flow of springs and seepage; dissolved and suspended mineral matter carried by ground and surface water and its effects on water use and quantities of water discharged by evaporation from lakes, streams, and the soil and vegetation. Hydrology is concerned with devising feasible methods for making these diverse measurements, accumulating and storing the data in usable form, and analyzing and interpreting them to solve practical water problems. Above all, it is concerned with the great task of making rigorous studies of all these basic data to determine principles and laws involved in the occurrence, movement, and work of the waters in the hydrologic cycle.

[A.N.S.]

Hydrolysis

Literally destruction, decomposition, or alteration of a chemical substance by water. In discussions of aqueous solutions of electrolytes the term hydrolysis is applied especially to reactions of cations (positive ions) with water to produce a weak base or of anions (negative ions) to produce a weak acid. A salt of a weak acid or of a weak base (or of both a weak acid and weak base) is then said to be hydrolyzed. The degree of hydrolysis is the fraction of the ion which reacts with the water. The term solvolysis is employed for reactions of solutes with solvents in general.

A frequently quoted example is the hydrolysis of sodium acetate, which is highly dissociated in water, to yield sodium ion, Na^+, and acetate ion, $CH_3CO_2^-$. The acetate ion is represented below by the symbol A^-. Because acetic acid is a weak acid (it has a small dissociation constant) the acetate ion, A^-, combines with some of the hydrogen ion present in aqueous solutions thus:

$$A^- + H_3O^+ \rightleftharpoons HA + H_2O \qquad (1)$$

The concentration of the hydrogen ion is involved in the following equilibrium between water and the hydrogen ion (often called the oxonium ion or hydronium ion) and the hydroxide ion:

$$2H_2O \rightleftharpoons H_3O^+ + OH^- \qquad (2)$$

Because reaction (1) removes a product of reaction (2), each reaction proceeds to produce more of its products (the substances on the right side of each equation). The net result is

$$A^- + H_2O \rightleftharpoons HA + OH^- \qquad (3)$$

Equation (3) represents the reaction which is described as the hydrolysis of the acetate ion. When equilibrium is attained the concentrations are such as to conform to the equilibrium constants of all three reactions (one of which is a mathematical consequence of the other two).

Textbooks contain a simple equation for the calculation of the degree of hydrolysis of an ion in a dilute salt solution.

For reaction (3), the degree of dissociation x is given by

$$x = \frac{[OH^-][HA]}{[A^-]} = \frac{[OH^-][H^+][HA]}{[H^+][A^-]} = \frac{K_w}{K_a}$$

where the brackets indicate molar concentrations, K_w is the ion-product constant of water, and K_a is the dissociation constant of the acid, HA.

As the equation implies, the fraction hydrolyzed may be affected in one of several ways.

1. The hydrolyzed fraction may become larger as the dissociation constant of the weak acid or base produced by hydrolysis becomes smaller. For example, sodium cyanide is more highly hydrolyzed than sodium acetate because hydrocyanic acid is a weaker acid than acetic acid; consequently a solution of sodium cyanide dissolved in pure water is slightly more alkaline than one of sodium acetate.

2. The hydrolyzed fraction may become larger as the dissociation constant of the solvent becomes larger. For example, the degree of hydrolysis of sodium acetate in water is larger at 50°C than at 25°C because the dissociation constant of water increases with temperature and because the dissociation constant of acetic acid is smaller at 50°C than at 25°C.

3. The fraction may become larger as the salt concentration becomes smaller. Because of oversimplification, the equation usually given in text-

books implies that the degree of hydrolysis approaches 100% as concentration approaches zero. Actually the limit approached is only about 0.5% for sodium acetate in carbon dioxide-free water at 25°C.

4. The fraction may become much larger if the electrolyte is the salt of both a weak acid and a weak base. If the dissociation constants are nearly equal at the temperature of the solution, the solution need be neither appreciably acid nor appreciably alkaline.

Other types of electrolytic hydrolysis may occur. Sometimes a solution may become sufficiently acidic or sufficiently alkaline to cause the formation by precipitation of another phase, for example, a solid basic salt. Even the anion of the strong acid HCl may be hydrolyzed if solid sodium chloride is heated in the presence of moisture. Some of the very volatile HCl is lost and the residual melt when cooled and dissolved in pure water is demonstrably alkaline. In this example the hydrolysis does not occur because of the formation of a weak acid in solution but because of the formation of a volatile compound. *See* ACID AND BASE; EQUILIBRIUM, IONIC; HYDROGEN ION; HYDROLYTIC PROCESSES.

[T.F.Y.]

Bibliography: W. C. Pierce, E. L. Haenisch, and D. T Sawyer, *Quantitative Analysis*, 4th ed., 1958.

Hydrolytic processes

Reactions of both organic and inorganic chemistry wherein water effects a double decomposition with another compound, hydrogen going to one component, hydroxyl to another.

$$XY + H_2O \rightarrow HY + XOH$$
$$KCN + H_2O \rightarrow HCN + KOH$$
$$C_5H_{11}Cl + H_2O \rightarrow HCl + C_5H_{11}OH$$

Although the word hydrolysis means decomposition by water, cases in which water brings about effective hydrolysis unaided are rare, and high temperatures and pressures are usually necessary. *See* HYDROLYSIS.

In the field of organic chemistry, the term hydrolysis has been extended to cover the numerous reactions in which an alkali is added to water, as in the hydrolysis of esters.

$$CH_3COOC_2H_5 + NaOH \rightarrow CH_3COONa + C_2H_5OH$$

Even more common is the application of the term to reactions in which *an acid* is added to water. The enormous production of glucose from starch exemplifies such reactions. The addition of acids and alkalies hastens such reactions even if it does not initiate it.

Hydrolytic reactions may be classified as follows: (1) hydrolysis with water alone; (2) hydrolysis with aqueous acid, *dilute* or concentrated; (3) hydrolysis with aqueous alkali, dilute or concentrated; and (4) alkali fusion with little or no water at high temperatures.

Water alone can be used to hydrolyze organometallic compounds such as Grignard reagent and zinc alkyls. It is also useful in hydrolyzing (or hydrating) acid anhydrides, lactones, and other internal anhydrides.

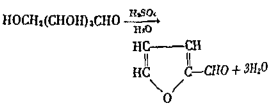
Material susceptible to hydrolysis. Saturated aliphatic hydrocarbons do not enter into hydrolytic reactions, even in the presence of acids or alkali. The olefins, however, react readily. The direct synthesis of ethanol from ethylene in the presence of a phosphoric acid catalyst is carried out on a large scale.

$$C_nH_{2n} + H_2O \rightarrow C_nH_{2n+1}OH$$

Ethanol and isopropyl alcohol can also be obtained by first treating the olefins with concentrated sulfuric acid and then hydrolyzing the resultant alkyl hydrogen sulfate with steam.

$$C_nH_{2n} + H_2SO_4 \rightarrow C_nH_{2n+1}OSO_3H \xrightarrow{H_2O} C_nH_{2n+1}OH + H_2SO_4$$

The hydrolysis of wood by dilute sulfuric or concentrated hydrochloric acids produces glucose. A major world industry is based on the conversion of starch into maltose and glucose by treatment with hydrochloric acid. Other carbohydrates such as pectin and xylan react similarly. The pentosans, such as xylose, yield furfural.



Ethyl ether can be hydrolyzed to ethanol in the presence of dilute (10%) sulfuric acid at 272°C and a pressure of over 225 psi. The secondary alcohols are formed more readily than primary. Ethylene oxide and trimethylene oxide are converted readily to ethylene and propylene glycols, respectively.

Organic halides differ markedly in their behavior toward hydrolyzing agents. Acid halides react readily with water alone. Alkyl halides such as ethyl and amyl chlorides require elevated temperatures to effect conversion.

$$C_5H_{11}Cl + NaOH \rightarrow C_5H_{11}OH + NaCl$$

The aromatic halides, such as chlorobenzene, are converted only at high temperatures (360°C) and high pressures (4200 psi).

Benzenesulfonic acid, when treated with steam under pressure, yields benzene and sulfuric acid,

$$C_6H_5SO_3H + H_2O \rightarrow C_6H_6 + H_2SO_4$$

but when fused with caustic soda, sodium phenoxide is the product.

$$C_6H_5SO_3H + 3NaOH \rightarrow C_6H_5ONa + Na_2SO_3 + 2H_2O$$

Equipment for hydrolysis. The hydrolytic production of such diverse products as fatty acids, furfural, ethanol, glycerol, phenol, and glucose, which

proceed by diverse mechanisms, requires specially designed equipment. Unlike nitrations or reductions, it is not feasible to design or describe any standard type of hydrolyzer. The design of hydrolyzing equipment and its materials of construction are related to specific operations.

Most hydrolytic reactions take place either in acid or alkaline media. The acids used are largely confined to sulfuric or hydrochloric; the bases include caustic soda or potash and the alkaline salts of weak acids. The acids may be dilute or concentrated; the bases, aqueous or fused.

Alkaline processes, with only minor exceptions, can be carried out in iron or steel equipment. Only in a few instances does caustic soda attack iron to produce hydrogen. Notwithstanding the severe treatment, the cast-iron pots used for caustic fusions (for example, of sodium benzenesulfonate) have an economically long life.

The use of sulfuric and hydrochloric acid in hydrolyses introduces special problems. In operations involving sulfuric acid, a high-silicon iron is usually a satisfactory construction material. A lead-lined vessel is useful for reactions involving dilute sulfuric acid.

Hydrochloric acid as a hydrolytic agent and as a product of hydrolysis introduces a materials problem because it is one of the most corrosive chemicals. Dry hydrogen chloride can be handled in iron equipment. Nickel, monel metal, and copper-base alloys are fairly resistant to dilute hydrochloric acid; tantalum equipment has outstanding resistance.

Typical products. Animal and vegetable oils and fats are glyceryl esters of fatty acids, similar and dissimilar, saturated and unsaturated. The hydrolysis of such fats with steam or acid produces glycerol and fatty acids. When sodium hydroxide (NaOH) is employed as the hydrolyzing agent, soap is produced.

$$\begin{array}{l} C_{17}H_{35}COO{-}CH_2 \\ \qquad\qquad\quad | \\ C_{17}H_{35}COO{-}CH \quad + 3NaOH \rightarrow \\ \qquad\qquad\quad | \\ C_{17}H_{35}COO{-}CH_2 \end{array}$$

Glyceryl tristearate

$$\begin{array}{lr} & CH_2OH \\ & | \quad \\ 3C_{17}H_{35}COONa + & CHOH \\ & | \quad \\ & CH_2OH \end{array}$$

Sodium stearate   Glycerol

Billions of pounds of glucose are produced annually by the hydrochloric acid hydrolysis of corn starch. These hydrolyzates are characterized as syrups, which do not crystallize readily on standing, and sugars, which are more completely saccharified and do crystallize on standing. The syrups have a dextrose equivalent (D.E.) of 24–60, the corn sugars have a D.E. of 82 and higher. The hydrolytic operations may be either batch or continuous.

The abundance and relative cheapness of the olefins from petroleum refinery operations ha[ve led] to their utilization for the production of alcohols. From the alkyl hydrogen sulfates obtained by the treatment of olefins with sulfuric acid, the following alcohols are now manufactured: ethyl from ethylene, isopropyl from propylene, secondary butyl from normal butylene, tertiary butyl from isobutylene, and secondary and tertiary amyl from pentenes. Ethanol is also made by the direct addition of water to ethylene by a vapor-phase process, using phosphoric acid on Celite as a catalyst.

$$C_2H_4 + H_2O \xrightarrow{H_3PO_4} C_2H_5OH + 19{,}000 \text{ Btu/lb-mole}$$

The production of synthetic phenol from benzene exemplifies three types of hydrolytic reaction:

$$C_6H_5Cl + NaOH \rightarrow C_6H_5OH + NaCl \quad (1)$$

$$C_6H_5SO_3Na + 2NaOH \rightarrow C_6H_5ONa + Na_2SO_3 \quad (2)$$

$$C_6H_6 + HCl + \tfrac{1}{2}O_2 \rightarrow C_6H_5Cl + H_2O \quad (3a)$$

$$C_6H_5Cl + H_2O \xrightarrow[\text{catalyst}]{\text{Phosphate}} C_6H_5OH + HCl \quad (3b)$$

All these processes are used commercially; (1) is a liquid-phase reaction under high pressures, (2) is an example of a fusion, and (3a) and (3b) are continuous vapor-phase reactions. See ALCOHOL; ALCOHOLYSIS; FAT AND OIL, NONEDIBLE; HALOGENATED HYDROCARBON; UNIT PROCESSES.

[P.H.G.]

Hydromechanics

That branch of physics that deals with liquids, traditionally water, as a medium for the transmission of forces (see HYDRAULICS; HYDROSTATICS). A body immersed in a liquid experiences a vertical force (see BUOYANCY) proportional to the volume of liquid that it displaces (see ARCHIMEDES' PRINCIPLE). A fluid whose velocity changes produces forces as a result of changing pressure (see BERNOULLI'S THEOREM) and of changing momentum (see WATER HAMMER). Such phenomena are applied in various machines (see HYDRAULIC PRESS; HYDRAULIC TURBINE).

[W.A.]

Hydrometallurgy

Processes employing chemical reactions in aqueous solution for the extraction of metal values from ores and concentrates. Hydrometallurgical processes are of three kinds: leaching, the selective dissolution of metal values from the ore into water solution to form the pregnant solution; extraction, the treatment of pregnant solution with a second phase (exchange resin or organic solvent) to cause selective extraction of solution constituents into the new phase; precipitation, removal of metal values from the pregnant solution by formation of an insoluble substance containing the metallic element.

The advantages of hydrometallurgy are: applicability to very low-grade ores (gold, uranium); adaptability to difficult separations of similar metals (hafnium from zirconium); absence of large fuel requirements and simplified materials handling compared with pyrometallurgy. With few exceptions hydrometallurgy produces purified metal co[m]pounds rather than metals, hence must be foll[owed]

by a process of pyrometallurgy or electrometallurgy to yield metal.

Modern hydrometallurgy dates from the introduction, in 1889, of the cyanide process for recovery of gold and silver from low-grade ores. The versatility of hydrometallurgy was greatly extended after World War II by development of pressure leaching, pressure precipitation, ion exchange, and solvent extraction. At mid-twentieth century, hydrometallurgy occupied an important role in the production of aluminum, magnesium, uranium, zinc, nickel, copper, cobalt, gold, silver, tungsten, molybdenum, and other metals. *See* ION EXCHANGE; LEACHING; METALLURGY; PRECIPITATION (CHEMISTRY); SOLVENT EXTRACTION. [H.H.K.]

Hydrometeorology

The study of the occurrence, movement, and changes in the state of water in the atmosphere. The term is also used in a more restricted sense, especially by hydrologists, to mean the study of the exchange of water between the atmosphere and continental surfaces. This includes the processes of precipitation and direct condensation, and of evaporation and transpiration from natural surfaces. Considerable emphasis is placed on the statistics of precipitation as a function of area and time for given locations or geographic regions.

Water occurs in the atmosphere primarily in vapor or gaseous form. The average amount of vapor present tends to decrease with increasing elevation and latitude and also varies strongly with season and type of surface. Precipitable water, the mass of vapor per unit area contained in a column of air extending from the surface of the earth to the outer extremity of the atmosphere, varies from almost zero in continental arctic air to about 8 g/cm^2 in very humid, tropical air. Its average value is about 3 g/cm^2, an amount equivalent to a column of liquid water slightly greater than 1 in. in depth.

Atmospheric water cycle. Although a trivial proportion of the water of the globe is found in the atmosphere at any one instant, the rate of exchange of water between the atmosphere and the continents and oceans is high, and the atmosphere is extremely mobile. The average water molecule remains in the atmosphere about 3 days and is precipitated many hundreds or even thousands of miles from the place at which it entered the atmosphere.

A major feature of the atmospheric water cycle is the net flux of water vapor meridionally, from one latitude to another. The average precipitation exceeds evaporation in a narrow band extending approximately from 5°S to 10°N lat. To balance this, the atmosphere carries water vapor equatorward in the tropics, primarily in the quasi-steady trade winds which have a component of motion equatorward in the moist layers near the earth's surface. P[illegible] the temp[illegible] spheres, [illegible] lat. In the middle and higher latitudes, therefore, the atmosphere carries vapor poleward. Here the exchange occurs through the action of cyclones and anticyclones, large-scale eddies of air with axes of spin normal to the earth's surface.

Table 1. Meridional flux of water vapor in the atmosphere

Latitude	Poleward flux, 10^{10} g/sec
90°N	0
70°N	7
40°N	59
10°N	−60
10°S	−40
40°S	91
70°S	3
90°S	0

For the globe as a whole the average amount of evaporation must balance the precipitation. The subtropics are therefore regions for which evaporation substantially exceeds precipitation. The complete meridional cycle of water vapor is summarized in Table 1. This exchange is related to the characteristics of the general circulation of the atmosphere. It seems likely that a similar cycle would be observed even if the earth were entirely covered by ocean.

Additional complications arise from the existence of land surfaces. Over the continents the only source of water is from precipitation; therefore, the average evapotranspiration (the sum of evaporation and transpiration) cannot exceed precipitation. The flux of vapor from the oceans to the continents through the atmosphere, and its ultimate return to atmosphere or ocean by evaporation, transpiration, or runoff is known as the hydrologic cycle (*see* HYDROLOGY). Its atmospheric phase is closely related to the air mass cycle. In middle latitudes of the Northern Hemisphere, for example, precipitation occurs primarily from maritime air masses moving northward and eastward across the continents. Statistically, precipitation from these air masses substantially exceeds evapotranspiration into them (*see* EVAPOTRANSPIRATION). Conversely, cold and dry air masses tend to move southward and eastward from the interior of the continents out over the oceans. Evapotranspiration into these continental air masses strongly exceeds precipitation, especially during winter months. These facts, together with the extreme mobility of the atmosphere and its associated water vapor, make it likely that only a small percentage of the water evaporated or transpired from a continental surface is reprecipitated over the same continent.

Precipitation. Hydrometeorology is particularly concerned with the measurement and analysis of precipitation data. In the United States alone more than 12,000 stations report precipitation daily, and approximately 3500 stations maintain recording rain gages. Since 1950, increasing attention has been paid to the use of radar in estimating precipitation. By relating the intensity of radar echo to rate of precipitation it has been possible to obtain a vast amount of detailed information concerning

the structure and areal distribution of storms. See METEOROLOGICAL INSTRUMENTATION.

Precipitation may, of course, be in liquid or solid form. In addition to rain and snow there are other forms which often occur, such as hail, snow pellets, sleet, and drizzle. Whatever the type, appreciable quantities of precipitation are necessarily associated with upward vertical motion of air. If this upward motion occurs uniformly over a wide area, measured in tens or hundreds of miles, the associated precipitation is usually of light or moderate intensity and may continue for a considerable period of time. Vertical velocities accompanying such stable precipitation are usually of the order of several centimeters per second. Under other types of meteorological conditions, upward velocities may locally be very large (of the order of several meters per second) and may be accompanied by compensating downdrafts. Such convective precipitation is best illustrated by the thunderstorm. Intensity of precipitation may be extremely high, but areal extent and local duration are comparatively limited. Storms are sometimes observed in which local convective regions are imbedded in a matrix of stable precipitation. There are four basic types of precipitation.

Orographic precipitation. This form occurs when warm, moist air is lifted by topographic barriers.

Frontal precipitation. The frontal type is somewhat similar to orographic precipitation, except that the precipitating air mass is forced aloft along a front above a colder, more dense air mass which itself may be in motion relative to the earth's surface (see FRONT).

Air mass precipitation. This is associated with vertical motion of warm, moist air brought about by converging horizontal winds in the absence of contrasting air masses.

Precipitation from tropical storms. Although technically a type of air mass precipitation, this variation is of such importance that it is usually considered separately.

Analysis of precipitation data. Precipitation is essentially a process which occurs over an area. However, despite the experimental use of radar, most observations are taken at individual stations. Analyses of such "point" precipitation data are most often concerned with the frequency of intense storms. These data are of particular importance in evaluating local flood hazard, and may be used in such diverse fields as the design of local hydraulic structures, such as culverts or storm sewers, or the analysis of soil erosion. Intense local precipitation of short duration (up to 1 hour) is usually associated with thunderstorms. Precipitation may be extremely heavy for a short period, but tends to decrease in intensity as longer intervals are considered. Several record point accumulations of rainfall are shown in Table 2.

Hydrometeorologists must often consider the statistics of occurrence or lack of occurrence of precipitation over an area, such as a watershed, or an economic or even a political region. The intensity of extreme storms tends to decrease with increasing area, as well as with increasing duration. A typical hydrometeorological problem might involve estimating the likelihood of occurrence of a storm of given intensity and duration over a specified watershed, for purposes of determining the required spillway capacity of a dam. Such estimates can only be obtained from a careful meteorological and statistical examination of large numbers of storms selected from climatological records. In the United States the U.S. Army Corps of Engineers, in cooperation with the Weather Bureau, has embarked on a continuing program of analysis to make such historical depth-area-duration data available to the practicing engineer.

When large hydraulic projects are undertaken, it may be desirable to design a structure to meet the worst conceivable conditions. In such cases it is necessary to estimate the maximum possible storm that can occur over the area in question. Obviously, there is no rigorous way in which this can be done, nor is an exact definition of the term itself possible. Nevertheless, by examining historical meteorological records and by postulating a combination of extremely moist air with a wind field calculated to produce an extreme storm, an estimate can be obtained which is useful for design purposes.

Evaporation and transpiration. In evaluating the water balance of the atmosphere, the hydrometeorologist must also examine the processes of evaporation and transpiration from various types of natural surfaces, such as open water, snow and ice fields, and land surfaces with and without vegetation. From the point of view of the meteorologist, the problem is one of transfer in the turbulent boundary layer. It is complicated by topographic

Table 2. Record observed point rainfalls

Duration	Depth, inches	Station	Date
1 min	0.65	Opid's Camp, Calif.	April 5, 1926
5 min	2.48	Porto Bello, Panama	Nov. 29, 1911
15 min	7.80	Plumb Point, Jamaica	May 12, 1916
42 min	12.00	Holt, Mo.	June 22, 1947
2 hr, 45 min	22.00	Near D'Hanis, Texas	May 31, 1935
24 hr	45.99	Baguio, Philippine Is.	July 14-15, 1911
1 month	366.14	Cherrapunji, India	July, 1861
12 month	1041.78	Cherrapunji, India	Aug., 1860 to July, 1861

effects when the natural surface is not homogeneous. In addition the simultaneous heating or cooling of the atmosphere from below has the effect of enhancing or inhibiting the transfer process. Although the problem has been attacked from the theoretical side, empirical relationships are at present of greatest practical utility. *See* METEOROLOGY. [C.S.B.]

Bibliography: T. F. Malone (ed.), *Compendium of Meteorology*, 1951.

Hydrometer

An instrument used to measure the specific gravity of a liquid. The hydrometer consists of a float with a small stem, so weighted that the float is partially immersed and the stem rises vertically from the surface of the liquid (Fig. 1). The float sinks in the liquid until the buoyant effect of the liquid on the immersed volume equals the weight of the float. With heavy (dense) liquids the emergent stem is long; with light liquids the emergent stem is short.

Fig. 1

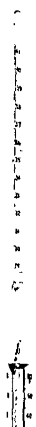


Fig 2

Fig 1. Plain hydrometer. (Taylor Instrument Co.)
Fig 2. Thermohydrometer. (Taylor Instrument Co.)

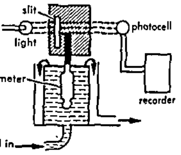


Fig. 3. Photoelectric hydrometer. *(From D. M. Considine, ed., Process Instruments and Controls Handbook, McGraw-Hill, 1957)*

The hydrometer is calibrated at a given temperature for use at that temperature. It may be calibrated in specific gravity with water at 4°C as a reference (1.000 on the scale), but normally water at 60°F is the reference. Alternatively, it may be calibrated in any one of a number of arbitrary scales which are characteristic of and have become standard in certain industries. The American Petroleum Institute (API) scale is used in the petroleum producing and refining industry. The Baumé scale is used primarily for acids and syrups. The Brix scale is used in the sugar industry, and the Plato (Balling) scale is used in the brewing industry. Others include the Quevenne, Barkometer, % alcohol, and Twaddle scales. Some hydrometers are calibrated for reading at the top of the meniscus, but most read correctly at the principal surface of the liquid Thermohydrometers (Fig. 2) incorporate a thermometer within the float so that the temperature and specific gravity of the liquid may be determined with a single instrument.

When hyd ... measure the specific gravity ... or similar t... when the liquid temperature and observation technique duplicates calibration conditions, hydrometers are accurate to 0.002 (sp gr). Certain constructions with large floats and small stems may be accurate to 0.0001, but instruments of this type are in limited use.

Modifications of the hydrometer are used in order to obtain continuous indication, recording, or control of a flowing liquid. A photoelectric hydrometer (Fig. 3) uses a light source and a photocell to measure changes in density. The inductance-bridge hydrometer (Fig. 4) and the same buoyant principle with a chain balance (Fig. 5) are also used. All these systems require sensitive low-drift amplifiers to obtain good density or specific gravity readings. The temperature of the liquid must be accurately controlled, or temperature compensation must be provided. In addition, an error is caused by the flow of fluid through the sample

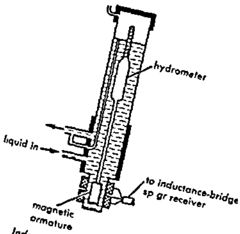


Fig. 4. Inductance-bridge hydrometer. (From D. M. Considine, ed., Process Instruments and Controls Handbook, McGraw-Hill, 1957)

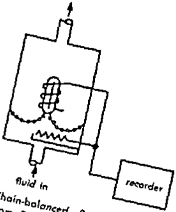


Fig. 5. Chain-balanced float-type density-sensing element. (From D. M. Considine, ed., Process Instruments and Controls Handbook, McGraw-Hill, 1957)

chamber, but this may be minimized at the expense of the dynamic response of the instrument by keeping the flow at a small value. *See* DENSITY MEASUREMENT; SPECIFIC GRAVITY. [R.E.CL.]

Bibliography: D. M. Considine (ed.), *Process Instruments and Controls Handbook*, 1957.

Hydrophone

A device which receives underwater sound waves and converts them to essentially equivalent electric waves. A hydrophone is the underwater analog of a microphone, the latter being an electroacoustic device which responds to sound waves in air (*see* MICROPHONE). Hydrophones are used in sonar apparatus, in sonobuoys, and in certain underwater weapons. For further discussion, *see* TRANSDUCER, UNDERWATER; *see also* ACOUSTIC MINE; ACOUSTIC TORPEDO; SONAR; SONOBUOY. [K.W.P.]

Hydroponics

One of several popular names given to the artificial culture of plants in inorganic salt solutions in open concrete tanks or other containers. The technique was devised in the mid-nineteenth century as a means of studying mineral nutrients essential to the growth and development of plants, but for many years was only a laboratory procedure.

Mineral solution culture. Many mineral salt combinations and ratios of concentrations have been investigated. No one composition of nutrient solution is superior to others in all cases. However, within certain ranges of composition and concentration, a fairly wide variety of nutrient solutions is suitable for the growth of most plants. Many problems of plant nutrition have been elucidated by employment of hydroponic techniques. The method simplifies such study since culture variables may be effectively applied and controlled. The composition of mineral nutrients supplied to the plant is known, and changes in the composition can be followed in the medium during plant development. The acidity of the culture can be measured and controlled. Osmotic effects of total solution concentrations on growth can be observed. Qualitative and quantitative determinations of nutrient requirement are simplified and accurately measurable in solution culture media. Also each

Growth of tomato plants in fertile soil, in nutrient solution, and in pure silica sand irrigated each day with nutrient solution. Fruit had been harvested for 7 weeks prior to taking the photograph. All plants made excellent growth and set large amounts of fruit in all three media. The general cultural conditions, such as spacing and staking, were the same. (California Agricultural Experiment Station, Circ. 347)

Table 1. A typical nutrient solution

Specific salts used		Concentration of individual elements		
Compounds	Single strength, μmoles/liter	Elements	Single strength, μg-atoms/liter	Parts per million
Macronutrient solution medium—m6*				
KNO_3	7000	N†	14,000	198
$Ca(NO_3)_2 \cdot 4H_2O$	3000	K	7,000	273
$MgSO_4 \cdot 7H_2O$	1000	Ca	3,000	120
$NH_4H_2PO_4$	1000	Mg	1,000	24
		S	1,000	32
		P	1,000	31
Micronutrient solution medium—m7				
KCl	100	Cl	100	3.54
H_3BO_3	25	B	25	0.27
$MnSO_4 \cdot H_2O$	5.0	Mn	5.0	0.274
$FeSO_4 \cdot 7H_2O$‡	4.0	Fe‡	4.0	0.22
$ZnSO_4 \cdot 7H_2O$	2.5	Zn	2.5	0.164
$CuSO_4 \cdot 5H_2O$	0.5	Cu	0.5	0.032
H_2MoO_4	0.1	Mo	0.1	0.0096

* [illegible]

[illegible] $FeSO_4$ is prepared as [illegible] liter culture medium

of the factors which alter or determine the absorption of mineral ions by roots can be more effectively investigated; these include temperature, aeration, pH, previous carbohydrate or salt status, respiratory inhibitors, and transpiration. Typical culture solutions useful for experimental work have been published (Table 1). To demonstrate essentialities of specific macronutrient or major elements, special salt combinations are necessary, one element required for growth being omitted in each case. For micronutrient or trace element deficiencies, the specific salt carrying the element in question is withheld.

Economic possibilities. Mineral solution cultures are fully adequate for maximal growth if the essentials for plant development are provided, including light, air, favorable temperature, and a support for the plants. The size of such cultures has been increased by adapting the technique to crop production. Means of root support have been provided by sand, gravel, and cinders; adequate aeration is maintained by pumps or in other ways. Under these conditions, it is possible to obtain excellent crops in places where conventional methods of agriculture or gardening are unsuitable or impossible (Tables 2, 3). Economic aspects of artificial culture, however, must be compared with soil cultivation in order to judge the feasibility of the soilless practice at a particular location. It is doubtful that nutriculture crop production will ever completely supplant the usual agriculture method. Where economically feasible, it can effectively supplement present means of plant culture in soil. For example, it has proven useful in locations isolated from common food supply

Table 2. Reported yields* by hydroponics and conventional methods in soil†

Crop	Yields		Location
	Agricultural	Hydroponic	
Tomato, lb/plant	12	16.2	United States
	11	16.4	United Kingdom
	10	22.5	India
Potatoes, tons/acre	30	65	United States
Rice, lb/acre	900	5,000	India
	3000	9,000	Italy, Japan
Maize, lb grain/acre	2000	6,000	Bengal
Lettuce, lb/acre	9000	21,000	India
Beetroot, lb/acre	9000	20,000	India

* Differences for a specific crop are probably related to poor soil conditions and lack of fertilization, or uncomparable conditions.

† Modified from UNESCO, *Impact of Science on Society*, VI, 1955.

Table 3. Locations where hydroponic technique has proven useful

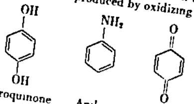
Location	Probable reasons for successful application
Iwo Jima and Wake Islands (Pacific Ocean) Aruba and Curaçao (Carribean Sea)	Lack of good agricultural soil, economics of shipping; water limitations
Habbaniyah (Iraq) and Bahrein (Persian Gulf)	Lack of good agricultural soil, water limitations
Bengal	Lack of good agricultural soil, economics of shipping; local [illegible]
Var[illegible]	[illegible]

sources and where the usual methods of crop growth are hampered by adverse conditions. It has also been used in certain areas where bacteria or other pathogens harmful to man enter plants from the soil. Crop quality and quantity of production are similar in nutriculture and in soil, provided conditions for growth are comparable. *See* FERTILIZING; PLANT, MINERALS ESSENTIAL TO. [T.C.B.]

Bibliography: See PLANT, MINERAL NUTRITION OF.

Hydroquinone

A dihydric phenol, also called quinol, in which the hydroxyl groups are para to each other on the ring of carbon atoms. It is produced by oxidizing aniline



Hydroquinone Aniline Quinone

to quinone with manganese dioxide and sulfuric acid, and then reducing the quinone to hydroquinone with iron and water. Its solutions become brown in air because of oxidation. This decomposition is very rapid in the presence of alkali. Hydroquinone is used to protect against deterioration caused by air oxidation and to prevent polymer-forming substances (monomers) from polymerizing prematurely. It is also used as a photographic reducer and developer. *See* PHENOL; QUINONE.

[R.B.C.]

Hydrosphere

Approximately 74% of the earth's surface is covered by water, in either the liquid or solid state. These waters, combined with minor contributions from ground waters, constitute the hydrosphere:

Oceans, including adjacent seas	1.4×10^9 km³
Continental ice	2.3×10^7 km³
Inland lakes	2.5×10^5 km³
Rivers and ground waters	2.5×10^5 km³

The oceans account for about 98% of the weight of the hydrosphere, while the amount of ice reflects the earth's climate, being higher during periods of glaciation. (Water vapor in the atmosphere amounts to 1.3×10^4 km³.)

The circulation of the waters of the hydrosphere results in the weathering of the land masses. Ocean water evaporates, forming rain, part of which falls on the continents. This water, partly taken up by the ground and partly by the streams, acts as an erosive agent before returning to the seas. *See* GROUND WATER; HYDROLOGY; SEA WATER.

[E.D.G.]

Hydrosphere, geochemistry of

The oceans of the world constitute a principal reservoir for substances in the major sedimentary cycle, which involves the processes of transport of material from the earth's crust to the sea floor. The cycle begins with the precipitation of water, acidified by the uptake of carbon dioxide in the atmosphere, onto the continents. This results in the physical and chemical breakdown of exposed surfaces. A part of the weathered material, in dissolved or solid states, is borne by the rivers to the oceans. Evaporation at the oceanic surfaces provides atmospheric water which precipitates in part upon the continents. This latter process completes the cycle. Table 1 gives the quantitative details by contrasting the marine areas and the respective land areas draining into the world's oceans. Clearly, per unit area, the Atlantic receives the weathering products from an integrated drainage area six times larger than that from the Pacific. The interior drainage areas are responsible for such water bodies as the Great Salt Lake, the Caspian Sea, and the Dead Sea.

Table 1. Oceanic and land-drainage areas, in 1000 km²

Ocean	Area	Land area drained	Ratio, area drained / ocean area
Atlantic	98,000	67,000	0.684
Indian	65,500	17,000	0.260
Antarctic	32,000	14,000	0.410
Pacific	165,000	18,000	0.110
Interior drainage		32,000	

Table 2. Residence times of elements in the oceans

Element	Amount in oceans, g	Residence time, years
Na	1.5×10^{22}	2.6×10^8
Ca	5.6×10^{20}	8.0×10^6
U	5.2×10^{15}	6.5×10^5
Cu	5.2×10^{15}	6.5×10^4
Si	5.2×10^{18}	1.0×10^4
Mn	1.4×10^{15}	7.0×10^3
Pb	$<4.0 \times 10^{14}$	$<1.0 \times 10^4$
Th	$<2.8 \times 10^{12}$	$<1.4 \times 10^3$
Ti	1.4×10^{18}	1.6×10^2
Al	1.4×10^{16}	1.0×10^2
Fe	1.4×10^{16}	1.4×10^2

Oceanic waters. The reactivities of chemical species in the oceans are reflected in the average times spent there before precipitation to the sea floor. Those elements with short residence time in the oceans engage more readily in chemical reactions that result in the formation of solid phases than those elements with long residence times.

Residence times. The calculations of residence times are based upon a simple reservoir model of the oceans whose chemical composition is assumed to be in steady state, that is, the amount of a given element introduced by the rivers per unit time is exactly compensated by that lost through sedimentation. An elemental residence time may be defined then by $t = A/(dA/dt)$ where A is the amount of the element in the oceans and (dA/dt) is the rate of introduction or the rate of removal of the element from the marine hydrosphere. Table 2 gives values for a representative group of elements.

The element of longest residence, sodium, has a residence time within an order of magnitude of the age of the oceans, several billion years. Similarly, the more abundant alkali and alkaline-earth metals all have residence times in the range of 10^6 to 10^8 years, resulting from the relative lack of reactivity of these elements in marine waters.

Elements which pass rapidly through the marine hydrosphere in the major sedimentary cycle, such as titanium, aluminum, and iron, not only enter the oceans in part as rapidly settling solids but also are reactants in the formation of the clay and ferromanganese minerals (for a discussion of the inorganic regulation of the composition *see* SEA WATER).

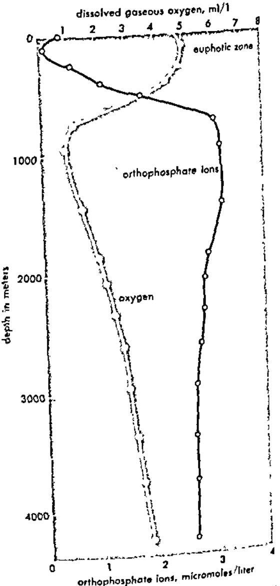
Although the relatively low values of these residence times are significant, the absolute values are probably unrealistic inasmuch as they are in conflict with an assumption used in their derivation. In treating the oceans as a simple reservoir, the mixing times of the oceans are assumed to be much greater than the residence times of the elements. Yet the oceans are believed to mix in times of the order of thousands of years. Nonetheless, it is significant that one may expect to find the concentration of such elements varying from one oceanic water mass to another.

Those elements with residence times of intermediate length, periods of the order of tens and hundreds of thousands of years, are probably of nearly uniform concentration in the oceans of the world, as are the chemical species of longer marine lives. Typical members of the group are such transition metals as copper, uranium, and nickel, elements in extremely dilute solution but actively involved in the inorganic chemistries of the seas. Such behavior is confirmed by the observation that these elements are in states of undersaturation in oceanic waters.

Photosynthesis. The presence of the large photosynthesizing biomass in the oceans gives rise to dramatic concentration changes. The amount of photosynthesis in the oceans, calculated to be of the order of 5×10^{17} tons of carbon dioxide consumed per year, compares with estimates for land of 7×10^{16} tons/year. The depth of the photosynthetic zone can extend downward from the surface to depths of 100 meters (m) or so depending upon the transparency of the water, season of the year, and latitude. In waters of active plant growth, carbon dioxide and oxygen (the intake and release gases of photosynthesis) are often observed in states of depletion and supersaturation, respectively.

The photosynthesizing plants require a group of dissolved chemical species, the nutrients, which are necessary for growth and multiplication. Ions of the orthophosphoric acids, nitrate, nitrite, and ammonia, as well as monomeric silicic acid, have very low concentrations in the regions of plant productivity as compared to regions of lesser fertility. Certain other substances concerned with plant growth, such as vitamins and trace metal ions, may have their marine concentrations governed by photosynthesis in surface waters, but as yet no definite relationships have been established.

The primary production of plant material furnishes the basis of nutrition for the animal domain of the oceans. The plant material which is removed from the photosynthetic zones but is not consumed by higher organisms, together with the organic debris resulting from the metabolic waste products or death of members of the marine biosphere, is oxidized principally in the oceanic waters of intermediate depth. This results not only in minimal values of dissolved gaseous oxygen at such depths but also in maximal concentrations of the nutrient



Distribution of dissolved gaseous oxygen and nutrient species of orthophosphate ions at 26°22.'4N and 168°57.'5W in the Pacific Ocean. (*Data from Chinook Expedition of 1956 of the Scripps Institution of Oceanography*)

species which are released subsequent to the combustion of the organic matter (see illustration).

The dissolved organic substances, arising from life in the sea, are of the order of 2–5 mg of carbon per liter. Higher amounts of particulate organic carbon are found in surface or coastal waters. *See* SEA WATER FERTILITY.

Salt content. For many problems in the physics of the sea and in engineering, the significant parameter is the total salt content, which governs the density as a function of pressure and temperature, rather than elemental, ionic, or molecular compositions.

The total salt content is expressed by either the salinity or the chlorinity, both terms given in units of parts per thousand (‰). The salinity is defined as the weight in grams in vacuo of solids which can be obtained from a water weight of 1 kg in vacuo under the following conditions: (1) the solids have been dried to a constant weight at 480°C; (2) the carbonates have been converted to oxides; (3) the organic matter has been oxidized; and (4) the bromine and iodine have been replaced by chlorine. A weight loss from the solid phases results from such chemistries, and hence the salinity of a given sample of sea water is somewhat less than its actual salt content.

In practice, the salt content is ascertained by the precipitation of the halogens with silver nitrate; the mass of halogens contained in 1 kg of sea water, assuming the bromine and iodine are replaced by chlorine, is designated as the chlorinity. The values so obtained are dependent upon the atomic weights of both silver and chlorine. Inasmuch as changes in the salt content of water are of interest when taken over long time periods, the chlorinity has been made independent of atomic weight changes by redefining it in terms of the weight of silver precipitated.

$$\mathrm{Cl}\ ‰ = 0.3285234\ \mathrm{Ag}$$

Chlorinity determinations are normally carried out with silver nitrate solutions which have been standardized against standard sea water, a water of known salinity which is obtainable from the Hydrographic Laboratory in Copenhagen, Denmark.

Chlorinity is related to the salinity.

$$\mathrm{S}\ ‰ = 0.030 + 1.8050\ \mathrm{Cl}\ ‰$$

Salinity of open ocean waters varies regionally between 32 and 37‰. Areas where evaporation exceeds precipitation, such as enclosed basins, are characterized by higher values. Salinities of 38–39‰ are representative of the Mediterranean Sea, while the northern part of the Red Sea has values ranging from 40 to 41‰. Coastal bays, subject to land drainage, and those waters which mix with meltwaters from cold regions, possess salinities which have all degrees of dilution comparable to those of the open ocean.

Although sea waters exhibit marked regional and depth differences in salt content, the ratios of the major dissolved constituents to one another, listed below, are almost invariable.

Constituents	*Ratio*	*Constituents*	*Ratio*
Na/Cl	0.555	S/Cl	0.0466
Mg/Cl	0.067	Br/Cl	0.0034
Ca/Cl	0.022	Sr/Cl	0.00042
K/Cl	0.0202	B/Cl	0.00024

Development of the hydrosphere. Many hypotheses on the origin and evolution of the earth's hydrosphere have been advanced over the past 50 years. Most of them can be placed in one of two categories: (1) the hypothesis of an original ocean which proposes that the present ocean has had much the same size and composition since the beginning of the geologic record; and (2) the hypothesis of continuous accumulation which considers that the ocean has been growing continuously, but not necessarily uniformly, since its inception.

Such considerations have certain common assumptions. First, the rock-forming species in sea water, sodium, potassium, silicon, iron, magnesium, and others, have been derived from the weathering of the earth's surface, whereas the marine quantities of water and the anionic constituents, such as chlorine and sulfur, cannot be adequately supplied by such a mechanism. These latter substances, as gases or dissolved species, have apparently evolved from the earth's interior by the degradation of the subsurface rocks. This proposal has been reached through the following arguments. The abundances of the noble gases, neon, nonradiogenic argon, and krypton, are many orders of magnitude less on the earth than in the universe relative to other elements. It is thus assumed that they were lost by the earth during its formative period. Therefore, substances which existed as gases and had comparable molecular weights were similarly not retained.

The hypothesis of the permanency of the oceans through geologic time complements the hypothesis of the permanence of the continents. Since old rocks are found in the basement of the central shield areas, with greater thicknesses of younger rocks at the edges, it has been assumed that the continents have grown laterally. Consequently, the reduction in area occupied by the oceans is compensated for by an increase in average depth. However, the calculated amounts of weathered materials that would form by the action of an initial ocean containing all the various anionic constituents as acids are far greater than those estimated by geologists to have been decomposed over all of geologic time. This hypothesis has been modified by some geochemists to one of an initial ocean of water with the gradual accretion of anionic substances.

The hypothesis of the slow growth of the oceans gains strength with the following observations. If only 1% of the hot-spring water is juvenile, the amounts found today, if extrapolated over geologic time, give a sufficient volume to produce the pres-

ent oceans. Further, the ratios of the major anionic constituents in sea water are similar to those found in plutonic gases.

Fresh waters and rain. Fresh continental waters show enormous variations in total salt content and in the relative concentrations of their various components, and such parameters in any given water body vary seasonally as well. Chemical analyses of a number of representative waters are given in Table 3.

The factors governing the composition of fresh-water bodies are many, and some are poorly understood. Significant amounts of the dissolved phases appear to come from rock weathering, organic decomposition products, atmospheric dusts, fuel combustion, air-borne particles from the marine environment, and volcanic emanations. Such materials, with the exception of the first, carried from their sources in the atmosphere, can be taken from the air to the river and lakes below by rain. It is not surprising then that the composition of rain is similar to that of diluted river waters. Table 3 indicates how the chemical analyses of typical rains approach those of fresh-water lakes.

Most fresh waters can be characterized as bicarbonate waters, with HCO_3^- exceeding all other anions, although in some instances chloride or sulfate is the dominant anion.

The sodium, chlorine, and often magnesium are dominantly of marine origin. They leave the oceans as sea spray and are air-borne to the continents. These elements show dramatic decreases in concentration in surface waters going from the coastal to the interior regions. Exceptions can be found in certain waters which derive their salts mainly by the denudation of igneous areas.

Calcium, and sometimes magnesium, can originate from drainage areas, as in the case of the Wisconsin fresh waters which drain over ancient magnesian limestones. Sulfate not only has sources in the marine environment but is also produced by the combustion of sulfur-containing fuels

Table 3. The chemical compositions of fresh waters and rain (values in parts per million)

Substance	Rivers*	Irish lakes†	English rain‡
Ca^{2+}	29 8	4 0	0 1–2 0
Na^+	8 4	8 6	0 2–7.5
Mg^{2+}	5 0	1 0	0 0–0 8
K^+	3 1	0 5	0 05–0 7
CO_3^{2-}	51 2	8 8	0 0–2 7
SO_4^{2-}	17 7	5 2	1 1–9 6
Cl	8 3	14 7	0 2–12 6
SiO_2	17 1		
NO_3	1 31		
PO_4^{3-}	0 34		
Fe_2O_3, Al_2O_3	4 02		

* E. J. Conway, Mean geochemical data in relation to oceanic evolution, *Proc. Roy. Irish Acad.*, 48B, 1942

† E. Gorham, The chemical composition of some western Irish fresh waters, *Proc. Roy. Irish Acad.*, 58B, 1957

‡ E. Gorham, On the acidity and salinity of rain, *Geochim. et Cosmochim. Acta*, 7, 1955

Table 4. The composition of waters from chloride, sulfate, and carbonate lakes*

Substance	Dead Sea	Little Manitou Lake	Pelican Lake, Oregon
Na^+	11.14	16 8	29 25
K^+	2.12	1.0	3 58
Mg^{2+}	13 62	10.9	2 62
Ca^{2+}	4 37	0 48	2 27
CO_3^{2-}	Trace	0 47	30 87
SO_4^{2-}	0 28	48.4	22 09
Cl^-	66 37	21.8	7.97
SiO_2	Trace	0 009	1 21
Al_2O_3, Fe_2O_3		0 21	0 02
Salinity	226,000	106,851	1983

* Data are from G. E. Hutchinson, *A Treatise on Limnology*, vol. 1, Wiley, 1957.

and from the oxidation of sulfur dioxide which results from the atmospheric burning of hydrogen sulfide, a product of the decomposition of organic matter.

The average salt content of fresh waters is of the order of 100 parts per million (ppm). Lower amounts of dissolved solids (50 ppm and less) are found in waters draining igneous rock beds, while open lakes and rivers carrying high salt contents (200 ppm and above) normally result either from the leaching of salt beds or from contamination by man.

Closed basins. The chemical compositions of closed basins, water bodies in which evaporation is the mechanism for the loss of water, are illustrated in Table 4, which contains representative examples of the three classical types, the carbonate, sulfate, and chloride waters. These classes appear in sequence during the removal of water from a system with the composition of average river or lake water. The carbonate types exist up until evaporation leads to the precipitation of calcium carbonate and to liquid phases enriched in sulfate and chloride ions. Further removal of water results in the precipitation of calcium carbonate and the subsequent precipitation of gypsum, $CaSO_4 \cdot 2H_2O$. The residual waters hence contain chloride as the dominant anion. *See* EVAPORITE (SALINE). [E.D.G.]

Bibliography: F. W. Clarke, *The Data of Geochemistry*, 5th ed., USGS Bull. 770, 1924; E. D. Goldberg, The processes regulating the composition of sea water, *J. Chem. Educ.*, 35:116–119, 1958; H. W. Harvey, *The Chemistry and Fertility of Sea Waters*, 1955; G. E. Hutchinson, *A Treatise on Limnology*, vol. 1, 1957.

Hydrostatics

The study of liquids at rest. In the absence of motion, there are no shear stresses; the internal state of stress at any point is determined by pressure alone. Hence, the pressure at a point is the same in all directions. Pressure acts normally to all boundary surfaces. For equilibrium under gravity, regardless of the shape of the containing vessel, the pressure is uniform over any horizontal cross

section. Pressure varies with height or depth in accordance with the relation $dp = w\,dz$ where w is the specific weight of the liquid (pounds per cubic foot), and z is the height or depth in feet measured positive downward.

Two different reference levels are used in measuring pressure. For many engineering purposes, gage pressure is used with pressure measured relative to atmospheric pressure as zero. For most scientific purposes, pressure is referred to true zero. Normal atmospheric pressure at sea level caused by the weight of the air above is approximately 14.7 lb per square inch absolute (psia). *See* PRESSURE MEASUREMENT.

Applications of hydrostatics. Storage tanks, underwater tunnels, gates for hydraulic structures, walls, dams, sheetpiling and bulkheads, pressure measuring devices, hydraulic presses, and other pressure-actuated systems are applications of hydrostatics.

Hydrostatic forces on immersed surfaces. Force F is the pressure p multiplied by the area A on which it acts. Its magnitude is in pounds and its direction is normal to the area. It is a vector quantity and may be broken into components usually taken horizontally and vertically. On all surfaces, plane or curved, forces acting on elementary areas are evaluated as $dF = p\,dA$.

On a plane surface all elementary force vectors are parallel and the total force is the sum of the elementary forces. Hence, the total force is the product of the total area and the average pressure acting on it. The average pressure is that at the centroid of the area The total force is independent of the orientation of the plane surface on which it acts as long as the depth to its centroid is the same.

Pressure volume. In calculating the total hydrostatic force acting on a plane surface, a solid is imagined with the plane surface area as its base and the fluid pressure at each point on the base erected there as an altitude. The total force is the volume of the solid. If the total force were imagined to act at a single point, the center of pressure of the surface, it would pass through the center of gravity of the solid and be normal to the surface.

On a nonplane surface, the elementary force vectors are not parallel. The summation of their horizontal and vertical components gives respectively the horizontal and vertical components of the total force. The total force is determined from these components by vector addition. The horizontal component of the total force is that which would be exerted on the vertical plane projected area of the nonplane surface. The vertical component of the total force is the weight of the liquid volume extending vertically from the nonplane surface up to the free surface of the liquid. As single forces, these components would pass through the centers of gravity of their respective volumes. When combined, they give the magnitude, line of action, and point of application of the total force on the nonplane surface.

Buoyant force. This is the force exerted vertically upward by a fluid on a body wholly or partly immersed in it. Its magnitude is equal to the weight of the fluid displaced by the body (*see* ARCHIMEDES' PRINCIPLE). This value is also the vertical component of the fluid pressure force acting upward against the bottom of the body minus the fluid pressure force component (if any) acting vertically downward against the top of the body. If this buoyant force equals the weight of the body, the body will remain at the given level. If it exceeds the weight of the body the latter will rise, and vice versa. The buoyant force as a single magnitude acts vertically upward through the center of buoyancy which is the center of gravity of the displaced fluid.

Stability. The stability of a wholly or partly immersed body is determined by the relative positions of its center of gravity G and center of buoyancy B. The position of G depends upon the distribution of the mass within the body; the position of B depends upon the shape of the submerged portion of the body. If G lies directly below B, the body will be stable; under an angular displacement, a righting moment will tend to restore the body to its original position.

A floating body, depending upon its shape, may be stable even if G lies above B. An angular rotation or heel will not change the volume of the displaced fluid but may change its shape and the lateral position of B so that a righting moment through B and G may exist to restore the body to its original position. The point of intersection of the vertical line through the displaced position of B with the line drawn through G and the original position of B is M. If M lies above G for a given angle of heel, a righting moment will exist; if M lies below G an overturning moment will arise to capsize the body. In most ships, for angles of heel up to 10–15°, M remains in a practically constant position—the metacenter.

Pressure-measuring instruments. Two types of instruments measure pressure—gages and manometers.

Gages. A metallic element such as a curved tube or a flexible diaphragm which deforms under liquid pressure is the usual sensing element of a gage. The deformation is changed mechanically or electrically into a calibrated dial reading. The Bourdon gage, used for measurement of static or slowly changing pressure, converts mechanically the deformation of a curved metal tube into a reading in pressure units. Each gage is designed to be accurate for a selected range of pressures but does not give accurate readings of short-time pressure fluctuations.

A pressure transducer converts deformation of a metallic diaphragm into an electric current differential, which is calibrated to read in pressure units. Each transducer is designed to be accurate for a selected range of pressures and reacts to fluctuations of microsecond duration. The cost is

high and accessory instrumentation elaborate; application is generally to the measurement of rapidly fluctuating pressures.

Manometers. Glass tubes in which the height of a liquid is a measure of the pressure being sought are called manometers. A simple manometer is open to the atmosphere at one end and connected to the pressure source at the other; it is called a piezometer if its liquid is the same as that in the tank, pipe, or other device to which it is connected. Measured above a selected datum, the height to the top of the liquid in the open-ended tube is the piezometer head. If the pressure source is a pipe, the height of the piezometer liquid level above the pipe center is the pressure head in the pipe. The pressure in pounds per square foot is obtained by multiplying this height in feet by the specific weight of the liquid.

A differential manometer measures the difference in pressure between two sources. The glass tube is usually an erect or inverted U partly filled with a liquid other than those liquids (or gases) of which the pressure difference is desired. One leg of the U is connected to one pressure source, the other leg to the other source. If the pressure sources are at the same elevation and contain the same liquid, the pressure difference is the vertical distance (displacement) between the tops of the manometer liquid in the two legs of the U multiplied by the difference between the specific weight of the manometer liquid and that of the displacing liquid. Measurement of small differences can be magnified by using a manometer liquid of low specific gravity; for large pressure differences mercury is commonly used.

Pressure transmission. Pressure applied to a confined liquid is transmitted with equal intensity throughout the liquid and by it to all surfaces of the confining vessel or piping. Hence, a small force applied to a small area of a confined liquid can create a large force against a large area. If the small and large areas are pistons the device may be a hydraulic press or jack. Because the transmitting liquid is practically incompressible and its volume virtually constant, the linear movement of the large piston will be to that of the small piston in inverse proportion to their areas. The principle of multiplying a force by means of liquid pressure applies also to hydraulic brakes, power steering, control systems, and the like; the actuating force may be a pump instead of a small piston. *See* HYDRAULICS; HYDROMECHANICS. [W.A.]

Hydroxide

One type of hydroxyl compound (OH-containing) whose solutions have a bitter taste, feel slippery to the touch, [illegible] is dissolved in water, an hydroxyl compound of the type *XOH* is formed. These molecules, XOH, can be ionized to give either OH^- ions or H^+ ions. The former are called bases and the latter acids. Ordinarily, when X is a metal, bases are obtained, for example, NaOH, and when X is a nonmetal, acids are obtained, for example, $S(OH)_6$, which loses two molecules of water and exists as H_2SO_4. Some insoluble oxides, Al_2O_3, for example, have both acid and basic properties. This class of compounds is said to be amphoteric.

Although only the alkali metal hydroxides are extensively soluble in water, the moderately soluble alkaline-earth metal hydroxides, and even quite insoluble oxides, such as Fe_2O_3, have basic properties and react with acids.

Some moderately soluble hydroxides, such as $Ba(OH)_2$ and $Ca(OH)_2$, are strong bases, that is, highly ionized. When solutions of heavy metal salts are made basic, insoluble hydroxides are precipitated. In reality these compounds often do not contain OH^- ions although they have basic properties. They are hydrates of one form or another.

The compounds containing more water than is required for the hydroxide formula, $Mg(OH)_2 \cdot xH_2O$, for example, are called hydrous hydroxides. Those containing less than the required water, for example, $Al_2O_3 \cdot H_2O \cdot xH_2O$, are called hydrous hydrates; and those that contain water not in any fixed amount are called hydrous oxides, for example, $Fe_2O_3 \cdot xH_2O$.

Hydroxides are very important compounds industrially, sodium hydroxide being the most important. It is used in making other chemicals, rayon, petroleum products, and soap.

Hydroxides are not the only compounds used as bases; others are phosphates, carbonates, and oxides. *See* ACID AND BASE; HYDRATE. [E.E.WR.]

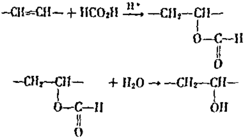
Hydroxylation reaction

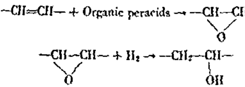
One of several types of reactions used to introduce one or more hydroxyl groups into organic compounds. Unsaturated compounds are the most common starting materials. Only those reactions having the widest use are discussed.

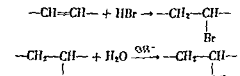
Monohydroxylation. Reaction (1) is especially useful when isomeric alcohols cannot form, as in the conversion of ethylene to ethyl alcohol. Reactions with longer chain olefins produce primarily secondary and branched-chain alcohols. Reaction (2) is relatively new, but promises to have wide use when the starting materials and products do not react with boiling formic acid. Lewis acids are preferred catalysts. Reaction (3) is of wide applicability. Hydrogenation can be effected catalytically or by reduction with metal hydrides. Reaction (4) is a classical one, and is rather limited.

$$-CH{=}CH- + H_2SO_4 \rightarrow -CH_2-\underset{\substack{|\\ SO_3OH}}{CH}- \qquad (1)$$

$$-CH_2-\underset{\substack{|\\ SO_3OH}}{CH}- + H_2O \rightarrow -CH_2-\underset{\substack{|\\ OH}}{CH}-$$

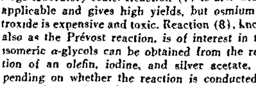

(2)

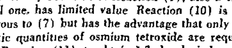

(3)

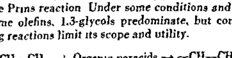

(4)

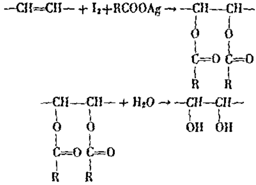
Dihydroxylation. Reaction (5) is the most widely used dihydroxylation reaction. Organic peracids need not be separately prepared, but most frequently are formed and utilized *in situ* by the interaction of hydrogen peroxide with formic or acetic acid. Reaction (6) is the classical dihydroxylation reaction, but the large amounts of manganese dioxide produced make the reaction awkward on a large laboratory scale. Reaction (7) is also widely applicable and gives high yields, but osmium tetroxide is expensive and toxic. Reaction (8), known also as the Prévost reaction, is of interest in that isomeric α-glycols can be obtained from the reaction of an olefin, iodine, and silver acetate, depending on whether the reaction is conducted in wet or dry acetic acid. Reaction (9), also a classical one, has limited value Reaction (10) is analogous to (7) but has the advantage that only catalytic quantities of osmium tetroxide are required.

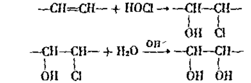
Reaction (11), to obtain 1,3-glycols, is known as the Prins reaction Under some conditions and with some olefins, 1,3-glycols predominate, but competing reactions limit its scope and utility.

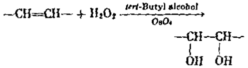

(5)

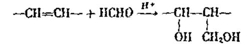

(6)


(7)


(8)


(9)


(10)


(11)

See ADDITION REACTION; EPOXIDATION; ETHYLENE GLYCOL; ETHYLENE OXIDE; GLYCOL; ORGANIC REACTION MECHANISM; SUBSTITUTION REACTION. [D.S.]

Hydrozoa

A class of the phylum *Coelenterata* which includes the fresh-water hydras, the marine hydroids, many of the smaller jellyfish, a few special corals, and the Portuguese man-of-war. The Hydrozoa may be divided into five orders—the Hydroida, *Milleporina*, Stylasterina, Trachylina, and Siphonophora. See separate article on each order.

Hydrozoa differ from the Scyphozoa, which are mostly the large jellyfish, and from the Anthozoa, to which the sea anemones and most corals belong, in the following features: there are no subdivisions of the digestive space by longitudinal partitions, nematocyst-bearing structures are lacking, and no stomodeum leads from the mouth. The medusa has a velum but lacks the highly specialized sense organs characteristic of the Scyphozoa.

The form of the body varies greatly among the hydrozoans. This diversity is due to the existence of two body types, the polyp and the medusa. A specimen may be a polyp, a medusa, a colony of polyps, or even a composite of the first two. Polyps are somewhat cylindrical, attached at one end, and have a mouth surrounded by tentacles at the free end. Medusae are free-swimming jellyfish with tentacles around the margin of the discoidal

In a representative life cycle, the fertilized egg develops into a swimming larva which soon attaches itself and transforms into a polyp. The polyp develops stolons which fasten to substrate, stems, and other polyps to comprise a colony of interconnected polyps. Medusae are produced by budding and liberated to feed, grow, and produce eggs and sperm. *See* METAGENESIS.

Hydroids conform to the foregoing description. Variation among them is due largely to differences in the pattern of growth and the extent to which the medusa stage is developed. In *Hydra* and a few hydroids, the medusa stage is absent, and new polyps, produced by budding, separate from the parent. In these cases each polyp is solitary (*see* HYDROIDA) Coral-like forms are similar to hydroids with the addition of a calcareous skeleton (*see* MILLEPORINA; STYLASTERINA). In some cases the polyp form is greatly reduced or absent and the medusa stage predominates as in the Trachylina. The greatest complexity occurs in the Siphonophora whose bodies have several different types of both medusoid and polypoid components.

Most hydrozoans are carnivorous and capture animals which come in contact with their tentacles. The prey is immobilized by poison injected by stinging capsules, the nematocysts Most animals of appropriate size can be captured, but small crustacea are probably the most common food. *See* COELENTERATA [S.CR.]

Bibliography: L. H. Hyman, *The Invertebrates*, vol. 1, 1940.

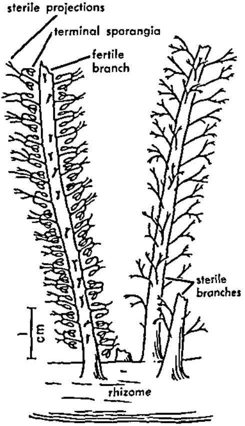
Hyeniales

An order of Devonian plants intermediate in evolution between the more primitive Psilophytales and more advanced orders of Sphenopsida Its small, dichotomously forked leaves tend to be whorled. The stem may be jointed Some leaves bear terminal sporangia, but these sporophylls are neither aggregated into a tight cone nor separated by bracts. *See* PSILOPHYTALES; SPHENOPSIDA.

Protohyenia from western Siberia, perhaps of Early Devonian ...

... axes borne on stout rhizomes. Its leaves are 2–4 times dichotomized (branched repeatedly into two branches) and their arrangement approaches the whorled condition. The tips of the fertile leaves are recurved and bear terminal sporangia. The sporophyll may have sterile projections beyond the sporangia.

Calamophyton, also from the Middle Devonian, consists of a stout axis that divides digitate-fashion into three or more lesser branches. These may dichotomize. The axes are frequently jointed and the whorled leaves are borne at the nodes. The leaves dichotomize 2–4 times. Some bear terminal sporangia on the recurved tips. The stem is siphonostelic (contains central pith) with a triangular



Hyenia, rhizome with fertile and sterile branches. (Redrawn from S. Leclercq)

primary xylem surrounding a triangular pith. Tracheids are chiefly scalariform. *See* PALEOBOTANY; PLANT KINGDOM; STELE. [H.P.B.]

Hygrometer

An instrument that directly measures relative humidity, or per cent moisture saturation of air. The moisture-sensing element in simple hygrometers is usually an organic material which expands and contracts with humidity changes. Hair, paper, membrane (goldbeater's skin), and wood are used most frequently. *See* HUMIDITY.

The sensitive material is placed under a small tension by a spring, and a magnifying linkage actuates a pointer as shown in Fig. 1. The effect of humidity on these elements varies with temperature; thus substantial errors exist at extreme values of humidity and temperature. Most instruments of this class are designed and calibrated for use at 70°F ± 20°F and relative humidity values between 15 and 90%. When the elements are exposed for extended periods to extremes of humidity or temperature (appreciably beyond the values specified above), they take a permanent set and are sometimes permanently damaged. The dynamic characteristics of hygrometers vary widely, but thin membrane, wood, and paper provide a reasonable time constant (possibly 5 min in a normal housing in air with natural convection).

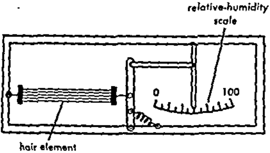


Fig. 1. Hair hygrometer. (From D. M Considine, ed., *Process Instruments and Controls Handbook*, McGraw-Hill, 1957)

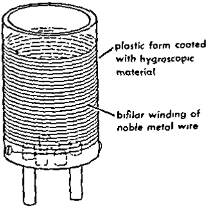


Fig. 2. Electrical hygrometer. (From D. M Considine, ed., *Process Instruments and Controls Handbook*, McGraw-Hill, 1957)

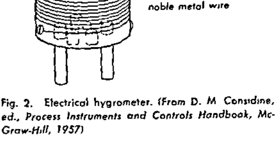


Fig. 3. Dew-point elements. (From D M. Considine, ed., *Process Instruments and Controls Handbook*, McGraw-Hill, 1957)

At best, a simple hygrometer should be regarded as an approximate instrument, unless it is checked periodically by a sling psychrometer under installed conditions. These same actuating elements, with more rugged designs, are used in industrial instrumentation for recording and control purposes. Normal maintenance procedures require periodic checking and zeroing of the instrument with the sling psychrometer. *See* PSYCHROMETER.

The electrical hygrometer measures the change in resistance in a hygroscopic material that is exposed to the air or gas. Inorganic salts, such as lithium chloride, are deposited on glass or plastic (Fig. 2), and the resistance changes due to humidity changes are measured by standard circuitry. The resistance-humidity curve is not linear and the element is temperature sensitive, so the unit is normally used with a calibration chart or over a limited range with circuit temperature compensation. A number of elements are normally required to cover the full humidity range (10–90%). Many other chemicals are hygroscopic and have been used in similar elements with limited success.

Another moisture-measuring instrument is illustrated in Fig. 3. The hygroscopic lithium chloride deposited on the insulating tape automatically carries enough current (ac) to maintain itself in vapor equilibrium with the moisture in the surrounding gas. The temperature in the element is measured by conventional means and bears a given relationship to the dew point of the air or gas. Its principal advantage is that it provides a continuous indication or record of dew point without the complication of conventional dew-point apparatus. *See* DEW POINT; MOISTURE-CONTENT MEASUREMENT.

[R.E.CL.]

Bibliography: D. M. Considine (ed.), *Process Instruments and Controls Handbook*, 1957.

Hymenoptera

The third largest order of insects, containing the sawflies, ants, wasps, bees, and related forms. A current catalog (1951) lists some 15,700 described species, subspecies, and varieties from America north of Mexico. Conservative estimates suggest that the world fauna may comprise well over 100,000 described species of this order, with many thousands still to be described.

It is an order of great importance to man. Some members, like the sawflies, certain chalcidoids, and most cynipoids, feed during the larval stage on foliage or other plant tissues Many species, such as the ichneumon flies, most chalcid flies, and wasps, are parasites or predators of other insects or spiders during their larval stage. Bees are indispensable in the pollination of many fruits, vegetables, and forage crops.

Hymenoptera occur in all major faunal zones but are more abundant and have greater diversity of species in the tropical and temperate zones Representation in the northern parts of the boreal zone is limited to a few sawflies, some Parasitica, and very few Aculeata, of which the bumblebees are the most conspicuous representatives (*see* INSECTA).

Diagnosis. Adult Hymenoptera usually may be recognized by having two pairs of membranous wings with reduced venation, the hind pair smaller than the front pair, and by mouthparts formed for biting, and often for lapping or sucking. In the higher forms, the abdomen is constricted basally, its first segment fused with the hind part of the thorax. Females always have an ovipositor modified for sawing, piercing, or stinging. Metamorphosis is complete, and sawfly larvae resemble

pillars except for the abdominal prolegs which lack a series of hooklets. Larvae of higher forms are legless and maggotlike but have a well-developed head. Pupae have the appendages free from the body except for certain chalcidoids.

The first four superfamilies of the Apocrita—the Ichneumonoidea, Chalcidoidea, Cynipoidea, and Proctotrupoidea—are commonly called the Parasitica, and the remaining superfamilies are known as the Aculeata. The Aculeata are stinging forms and the Parasitica are parasites of other insects. It is impossible to demarcate these two groups sharply because some Aculeata are parasites and some Parasitica are phytophagous. However, except for the phytophagous species of Parasitica, these insects lay their eggs in or on an insect or spider host, while the Aculeata place theirs in nests with a provision of food. Some connecting forms apparently occur in the Proctotrupoidea and Bethyloidea. Some of the Parasitica are more highly specialized, morphologically, than any other Hymenoptera, but the Aculeata show the greatest specialization in behavior. The table presents the major classification of the order as recognized in North America. The number of described species, subspecies, and varieties in America north of Mexico is noted for each family, and common names are given.

MORPHOLOGY OF HYMENOPTERA

The adult hymenopteron has a clearly differentiated head, thorax, and abdomen. Wings, when present, and legs are attached to the thorax.

Table 1. Families of Hymenoptera

Classification	Common Name	No of species
Suborder Symphyta	Sawflies	1009
Superfamily Megalodontoidea		120
Xyelidae		33
Pamphiliidae	Web-spinning sawflies	87
Superfamily Tenthredinoidea		849
Pergidae		13
Argidae		32
Cimbicidae	Cimbicid sawflies	12
Diprionidae	Conifer sawflies	35
Tenthredinidae	Sawflies	757
Superfamily Siricoidea		28
Syntexidae		1
Siricidae	Horntails	15
Xiphydriidae		6
Orussidae		6
Superfamily Cephoidea		12
Cephidae	Stem sawflies	12
Suborder Apocrita		13,346
Superfamily Ichneumonoidea		3814
Stephanidae		7
Braconidae	Braconid wasps	1239
Ichneumonidae	Ichneumon flies	2568
Superfamily Chalcidoidea		2032
Mymaridae	Fairy flies	110
Trichogrammatidae	Minute egg parasites	39
Eulophidae		544
Elasmidae		17
Thysanidae		18
Eutrichosomatidae		2
Tanaostigmatidae		4
Encyrtidae		320
Eupelmidae		89
Eucharitidae		27
Perilampidae		31
Agaontidae	Fig insects	2
Torymidae	Torymids	181
Ormyridae		17
Pteromalidae		321
Eurytomidae	Seed and stem chalcids	203
Chalcididae	Chalcids	101
Leucospidae		6
Superfamily Cynipoidea	Gall wasps	877
Ibaliidae		6
Liopteridae		2
Figitidae		58
Cynipidae	Cynipids or gall wasps	811
Superfamily Proctotrupoidea		985
Evaniidae	Ensign flies	11
Gasteruptiidae		50
Pelecinidae	Pelecinid wasps	1
Vanhorniidae		1
Roproniidae		3
Heloridae		1
Proctotrupidae		54
Ceraphronidae		101
Diapriidae		304
Scelionidae	Scelionid wasps	272
Platygasteridae		182
Trigonalidae		5
Superfamily Bethyloidea		345
Chrysididae	Cuckoo wasps	124
Bethylidae		100
Sclerogibbidae		1
Dryinidae		120
Superfamily Scolioidea		643
Tiphiidae	Tiphiid wasps	185
Sierolomorphidae		2
Mutillidae	Velvet ants	409
Rhopalosomatidae		2
Scoliidae		26
Sapygidae		19
Superfamily Formicoidea		786
Formicidae	Ants	786
Superfamily Vespoidea		368
Vespidae	Hornets, yellow jackets, potter wasps	368
Superfamily Pompiloidea		279
Pompilidae	Spider wasps	279
Superfamily Sphecoidea	Fossorial wasps	1215
Ampulicidae		3
Sphecidae		1212
Superfamily Apoidea	Bees	3304
Colletidae	Colletid bees	149
Andrenidae	Andrenid bees	852
Halictidae	Halictid and sweat bees	472
Melittidae		31
Megachilidae	Leafcutting bees	730
Apidae	Honeybees, bumblebees, carpenter bees	1070

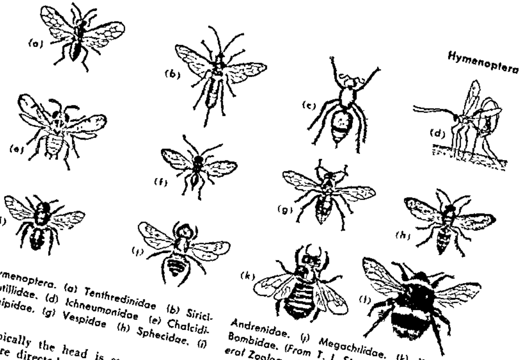


Fig 1. Hymenoptera. (a) Tenthredinidae (b) Siricidae. (c) Mutillidae. (d) Ichneumonidae (e) Chalcididae. (f) Cynipidae. (g) Vespidae (h) Sphecidae. (i) Andrenidae. (j) Megachilidae. (k) Xylocopidae. (l) Bombidae. (From T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957)

Head. Typically the head is so oriented that mouthparts are directed downward; however, all variations occur, and in some species the mouthparts are directed forward. The large compound eyes occupy much of the sides of the head, though they are reduced in size in many ants and some Parasitica. Three ocelli are typically present on the top of the head, but they may be reduced or absent in wingless forms. The paired antennae arise from the face between the eyes, and they may be close to the mouthparts or removed from them. Primitively, the antennae are long, slender, and multisegmented, but in the more highly specialized members of all major groups there has been a tendency toward reduction in the number of segments.

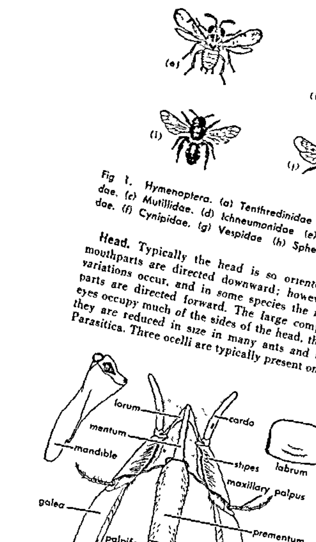


Fig. 2. Mouthparts of Anthophora edwardsii Cresson. (From E. O. Essig, College Entomology, Macmillan, 1942)

The mouthparts consist of paired mandibles, and a labiomaxillary complex, formed by membranous connections between the maxillae and labium. Most sawflies have simple biting mouthparts, but in the higher forms the labiomaxillary complex is variously modified to permit lapping or sucking types of feeding. In many bees and a few wasps, the labiomaxillary complex is greatly elongated to permit these insects to suck nectar from flowers which have deeply seated nectaries. The mandibles, in forms other than sawflies, do not function primarily in feeding but are used in manipulation of parasites or prey, construction of nests, and to escape from the cocoon or host body.

Thorax. The thorax consists of three segments, tightly fused together. Each segment bears a pair of legs, and each of the last two segments bears a pair of wings. In the sawflies, the thorax is broadly joined to the abdomen, but in the Apocrita the true first abdominal segment, the propodeum, has fused firmly with the thorax and is separated from the remainder of the abdomen by a constriction.

Wings. Most species have two pairs of wings, of which the posterior pair is the smaller. In flight, the fore- and hindwings are joined by a row of tiny hooks, the hamuli, along the fore margin of the posterior wing, which fit into the downfolded hind margin of the anterior wing. The most complete venation is shown by the sawflies, and extensive reduction of venation in varying degrees has occu

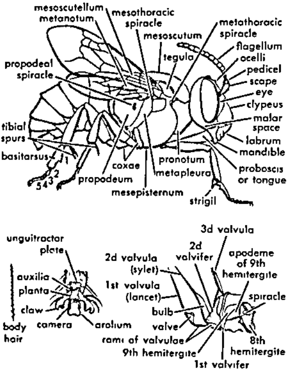


Fig. 3. The bee, *Anthophora edwardsii* Cresson. Drawings of adult male and the sting and accessory organs of the female to show the important characters used in classifying bees. (*From E. O. Essig, College Entomology, Macmillan, 1942*)

in all higher forms. Flightless species with shortened, nonfunctional wings, or no wings at all, occur in most major groups except the sawflies and bees. The females usually exhibit a greater reduction than do the males.

Legs. Each leg consists of a coxa which articulates with the corresponding thoracic segment, a trochanter, femur, tibia, and segmented tarsus. The tarsus usually has five segments, but these may be reduced to four or three in many Chalcidoidea and males of *Orussus* (Orussidae). The last tarsal segment typically bears a pair of claws. In most Parasitica, there is a constriction near the base of the femur which causes the trochanter to appear 2-segmented.

The legs are frequently modified to serve varied specialized uses. Many species which dig through the soil, like the Scolioidea, have thickened femora and tibiae bearing numerous stout spines; while other digging wasps have a comb of bristles on each fore tarsus which aids in raking the soil out of the burrow. Some bees, like *Andrena*, have groups of curled, branched hairs on the hindlegs to collect pollen, while others, like the honeybee and bumblebees, have the hind tibia expanded, flattened, and fringed by long curved hairs to form a pollen basket, or corbicula. Females of most Dryinidae have the fore tarsus modified into a chela to grasp the prey.

Abdomen. The abdomen primitively consists of 10 segments, though the number appears to be less because of modification or loss in the higher forms. In the sawflies, 10 terga and 9 sterna may be recognized, the apical parts being modified in connection with the male genitalia or female ovipositor. In the Apocrita, the first abdominal segment has fused with the thorax, and 6 other segments are normally visible in the female, 7 in the male. An exception to this is the Chalcidoidea, where there are 7 terga. Frequently there are fewer visible segments, either because of fusion, as in some Braconidae and Proctotrupidae, or because of retraction of the posterior 2 or 3 segments, as in the Chrysididae.

The female ovipositor, or sting, is formed from processes of the eighth and ninth sterna. In the sawflies, parts of the ovipositor have ridges which terminate ventrally in the teeth of the saw. Vestiges of such ridges in the Apocrita constitute the barbs, which, if well-developed as in the honeybee, cause the sting to remain in the wound. The eggs of both Symphyta and Parasitica pass through the ovipositor during oviposition. In the Aculeata, the ovipositor is purely a stinging mechanism, and the egg issues from an opening at its base. The sting is reduced or lacking in some ants and in the stingless honeybees of the Tropics.

Venom. In the Apocrita there is a pair of acid glands opening into a poison sac connected with the ovipositor. The secretion of these glands produces either a temporary paralysis when injected into their hosts by some Parasitica, or, usually, permanent paralysis when injected into their prey by aculeate wasps. Bees use their stings purely for defense. The venom of the Aculeata is a complex substance consisting of a protein and certain enzymes, as well as other constituents. Apparently the composition varies slightly with each species, which complicates the preparation of a desensitizing agent. When a human is stung, the enzymes react with his tissues to release histamine. Deaths may occasionally occur from anaphylactic shock, or from mechanical suffocation due to swelling of the lymphatic system. Medical assistance should be sought if severe swelling occurs following a sting, especially one on the face or throat.

Life history. Hymenoptera exhibit complete metamorphosis during development and pass through an egg, larval, and pupal stage. So far as is known, Hymenoptera always lay eggs. These are deposited in a protected situation on or near the supply of larval food. The hymenopterous egg is usually ovoid or sausage-shaped, and many species in some groups, like the Cynipoidea and some of the other Parasitica, have stalked eggs. The surface is usually unsculptured and very delicate. The eggs hatch after an incubation of variable length, and the resulting larvae begin a growth period consisting of three or four instars.

The larvae of sawflies are very similar in gross appearance to lepidopterous caterpillars. There is a well-developed head bearing powerful biting mouthparts. There are three pairs of segmented thoracic legs and, in all but the forms that bore in

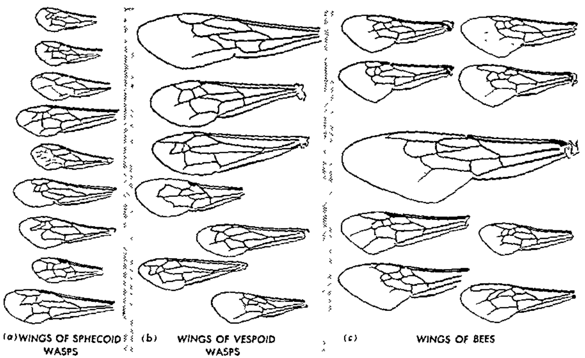


Fig. 4. Wings of certain Hymenoptera. (a) Wings of sphecoid wasps. (b) Wings of vespoid wasps. (c) Wings of bees. (From E. O. Essig, College Entomology, Macmillan, 1942)

stems or wood, there are from six to eight pairs of unsegmented abdominal prolegs which do not bear tiny hooks, or crochets, as in the Lepidoptera.

The larvae of Apocrita are legless, maggotlike forms in the later instars with a well-developed head and biting mouthparts. Occasionally, fleshy processes are present which aid in locomotion. The first instar larvae of some Apocrita, particularly those of the Parasitica and of the social parasites among the Aculeata, are very different from the larvae of later instars, a condition known as hypermetamorphosis.

The insect then spins a cocoon, if characteristic of the particular species, and enters the prepupal stage which may be of short or lengthy duration. The pupal stage usually lasts only two or three weeks, and the adult then ecloses, or emerges from the pupal case. The hymenopterous pupa has the appendages free from the body, except in certain Chalcidoidea. It is usually enclosed in a cocoon either constructed of silk alone or with other interwoven substances. The cocoon may be well developed, or it may be just a thin silken lining of the larval cell. It is absent in almost all Chalcidoidea and Cynipoidea, and, occasionally, in all other superfamilies. The adult usually remains in the cocoon for one to several days until the integument is thoroughly hardened, though occasionally the adult, as in some species of *Osmia* in the Megachilidae, may remain in the cocoon for ten months before beginning its active adult existence.

Some species have only one generation a year in temperate zones, others have two, and many breed continually during the warmer months. Hymenoptera usually overwinter as prepupae but occasionally ants, social wasps, and some bees overwinter as adults or as larvae, as in some Parasitica. So far as is known, none passes the winter in the egg or pupal stage.

BIOLOGY OF HYMENOPTERA

Practically all hymenopterous adults are terrestrial forms, living in, on, or near the earth's surface. A few species are secondarily aquatic, the adults swimming or walking under water to search out and parasitize aquatic or subaquatic hosts.

Most adults feed on plant nectar or honeydew secretions of various insects. A few sawflies prey on other insects. Some species of Parasitica and Aculeata imbibe body juices of the host or prey which they attack primarily for oviposition. Not too much is known of the nutritional requirements of newly emerged adults, but it is likely that many of them require some nutrient materials in order for the eggs to mature.

Reproduction. Mating takes place in a variety of situations, but it is always of rather short duration. Customarily, the males emerge from one to several days earlier than the females. Among the aculeates, the males of ground-nesting species may indulge in prenuptial flights over the nesting site while awaiting the emergence of females. Usually they pounce on the females and mating takes place on the ground. Males of some wood-nesting species may hover in front of the burrows harboring the females, and in this instance mating usually takes place during flight. Other species meet and mate on flowers. Among the ants and social wasps, large

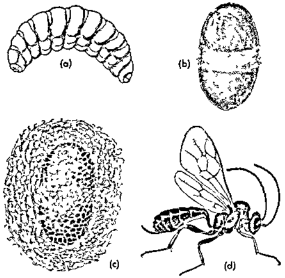


Fig. 5. Life cycle of *Bathyplectes curculionis* (Thomson), the parasite of the alfalfa weevil. (*a*) Larva. (b) Pupa. (c) Pupa within cocoon of the weevil. (*d*) Adult female (*From E. O. Essig, College Entomology, Macmillan, 1942*)

numbers of reproductive females and males emerge simultaneously from several colonies and form mating swarms in the air The honeybee has a similar nuptial flight, but it is composed of only one virgin queen and a number of drones (*see* SOCIAL INSECTS).

Most species are represented by both males and females. Males are usually produced from unfertilized eggs and have half the normal number of chromosomes, while females are produced from fertilized eggs and have the normal number of chromosomes. However, both facultative and obligate parthenogenesis are more common than in any other order of insects. In those species in which facultative parthenogenesis occurs, apparently only males are produced from virgin females. In those species exhibiting obligate parth- [illegible] female produces both sexes parthenogenetically, and the fertilized females of that generation produce only females.

In addition to males and fertile females found in most Hymenoptera, the social species have a third form called the worker caste. These are females which are sterile except under unusual circumstances. The workers of social wasps and bees are normally very similar in appearance to the queens, differing principally in their smaller size and occasionally in color pattern. The honeybee queen is quite different from the workers because she lacks wax glands and pollen-collecting apparatus, and she has a smooth rather than a barbed sting. Worker ants ordinarily differ more from their queens than do other species of social Hymenoptera —wings are absent, and the thorax is more modified. In some ants, the workers may be dimorphic, some of them being modified into a soldier form with huge head and mandibles for defense of the colony.

Dimorphism. Sexual dimorphism is often very marked. The two sexes of some species are so dissimilar that earlier students placed them in different genera or families. Even today there are many puzzles, and sexes have not been associated for many of the species having wingless females and winged males. Ordinarily the males are somewhat smaller than females, though the reverse is true in most species having wingless females (*see* SEXUAL DIMORPHISM).

In addition to sexual dimorphism, there are other kinds of variation in adult Hymenoptera. Species of Parasitica or of the social parasites among the Aculeata sometimes vary a great deal in size, depending on the size of the host or on the number of parasites per host. Also, there may be variation in form and size between the generations of species having more than one generation a year.

Behavior. Behavior of the Parasitica is much less complex than in the Aculeata. The primitive pattern is for the female to hunt for her host and then oviposit in or on it, without causing it any great disturbance. The host continues to live a normal life until it is killed by the developing parasite larva. This pattern is typical of many Ichneumonidae. A more complicated pattern is developed by stinging of the host to produce a temporary or permanent paralysis, or even death, before oviposition. This method is typical of many Braconidae and Chalcidoidea. Some species of Parasitica are hyperparasites or secondary parasites, that is, they are parasitic on parasites of the host. Some species may be either primary or secondary parasites, while tertiary parasitism, that is, species parasitic on secondary parasites, also has been reported in a few chalcidoids.

Nests and nesting behavior. The higher Aculeata are remarkable for the amount of care taken in construction of nests for their progeny. The nest may consist of a burrow in the ground or a boring in wood. The nest may also be constructed of clay, paper formed from masticated wood fibers, gums, resins, masticated leaf pulp, pieces of leaves, or wax made in the body of the insect. There may be only one cell per nest, a series of a dozen or so cells arranged in a clump or in a linear series, or there may be a multicellular nest, as in the social wasps and bees.

There are several basic types of nesting behavior among the aculeates. The most primitive pattern is for the wasp to search out its prey, paralyze it by stinging, and then to lay one or several eggs on it, leaving the prey where it was found. This is typical of most Bethylidae and many Scolioidea. A somewhat more complicated pattern, typical of most Pompilidae, is for the wasp to capture and paralyze its prey before constructing a nest for it in a sheltered situation. Another pattern is for the cell to be constructed and an egg deposited in it before the first specimen of prey is brought in. This occurs

typically among the Vespidae and a few of the Sphecidae which practice progressive provisioning. A fourth pattern is a variation of the third and consists of preparation of the cell, storing the cell with prey or a pollen-nectar mixture, and then ovipositiion. This method is typical of most Sphecidae and the Apoidea.

The solitary wasps store varied amounts of prey per cell. The Scolioidea, Pompilidae, and a few Sphecidae use only one specimen of prey per cell. Others, like the Vespidae and most Sphecidae, store from two to more than two dozen specimens of prey per cell depending upon the size of the prey used. Some solitary wasps are quite restrictive in their prey preferences and store only one species. The majority are not so selective and prey upon different species. However, most of these latter prey on species belonging to a single family or order, or to species occurring in a very restricted ecological niche. Thus, in the Sphecidae, species of *Pemphredon* prey on a number of species of aphids (Aphidae), and species of *Ectemnius* on a wide variety of Diptera, but *Symmorphus canadensis* in the Vespidae preys on leaf-mining larvae belonging to either the Coleoptera or Lepidoptera.

Sociality. Social parasites, also known as cuckoo wasps and bees, have arisen independently in several different families of wasps and bees. Except in the Chrysididae, ordinarily each is closely related, taxonomically, to the particular host wasp or bee which it parasitizes. These parasites either oviposit on the prey of the host wasp as the latter transports it to a nesting site, or they keep the nest under observation and slip into it at some time during the provisioning cycle to deposit the egg while the host female is absent from the nest.

Social life has arisen independently in several stocks of aculeate Hymenoptera, such as the ants, vespid wasps, halictid bees, bumblebees, stingless honeybees, and true honeybees. In the more primitive social insects, the colony dies out at the end of the year except for the recently emerged, fertilized queens which hibernate. These colonies are necessarily less populous than in the more advanced social forms, and are exemplified by the hornets and allies, the halictid bees, and bumblebees. In the higher social forms consisting of the ants, social wasps of the tropics, stingless honeybees, and true honeybees, the colonies are persistent and frequently number several thousands of individuals. The colonies of these wasps and bees divide by swarming, a process by which a queen and a number of workers leave the parent nest and begin a new colony.

Communication. One of the most remarkable facts that has been learned about the honeybee is that the bees employ a sign language to tell other members of the colony that they have located flowers with a copious flow of nectar or abundance of pollen, and approximately how far and in what direction this source is located. If the source is more than 100 yards from the hive, the worker performs a tail-wagging dance, usually on the vertical surface of a comb in the darkness. This dance consists of running over a figure-eight pattern with the part between the two loops of the 8 straight. The bee wags her abdomen from side to side when on this straight section. The number of runs over the straight section indicates the distance from the food source, and varies from nine to ten runs within 15 seconds for a distance of 130 yards, to three within 15 seconds for a distance of 830 yards. The direction of the source is indicated by the angle of the straight section of the dance. If straight up, the source is toward the sun; if straight downward, the source is directly away from the sun; if at an angle to one side or the other of the vertical, then the source is at a corresponding angle and direction away from the sun. If the food source is less than 100 yards from the hive, the worker performs a rapid round dance. The other bees do not receive any clue as to the direction or distance but only an indication that there is a copious food source nearby.

Reproductive behavior. Sawflies deposit their eggs on or in foliage, fruits, stems, or woody tissue, in accordance with the feeding requirements of the larva. Many Parasitica and a few Aculeata deposit more than one egg per host. Some species of Parasitica are polyembryonic, that is, from two to a thousand or more embryos may develop from a single egg. Usually the resulting progeny are all of one sex but exceptions occur. Polyembryony occurs in a few Braconidae, some species of *Macrocentrus*, some species of several genera of Encyrtidae, a few Platygasteridae (some *Platygaster*), and one species of Dryinidae, *Aphelopus theliae*.

The size and number of eggs laid is subject to great variation within the order. Most Symphyta and Chalcidoidea deposit a rather small number, from 10 to 50. Most of the Ichneumonoidea deposit larger numbers of eggs, from several hundred to more than a thousand. Some subsocial wasps which practice progressive provisioning may deposit only 6 eggs. Most aculeates lay from 10 to 75 eggs, but apparently some bethylids can lay 150 or more. Honeybees are probably the most fecund of the Aculeata. A queen may lay as many as 1500 eggs per day and 200,000 a year for at least 3 years.

Eggs of the Parasitica are deposited more or less at random on or in the host, but in the Aculeata each species deposits its egg at a particular location on the prey or in the nest.

Sawfly larvae are phytophagous and live on or in foliage, in stems, or in woody tissue. Larvae of the Parasitica may be either internal or external parasites, depending on whether the development takes place inside or outside the host. They can also be classified as solitary if there is only one larva per host, as in most Ichneumonidae, or gregarious, if there are two or more per host as in many Braconidae and Chalcidoidea. Larvae of most wasps are predaceous, feeding on paralyzed insects or spiders stored for them by the mother. Bee larvae feed on pollen-nectar mixture stored by the mother except for the social parasites which, in their

instar, are predaceous on the host egg or young larva and feed on the pollen-nectar mass in the later instars. The larvae of most Aculeata have a blind stomach during most of the larval life so that waste matter is not excreted until the end of the larval life when the connection between stomach and intestine is opened. The larvae of a few wasps, like Pemphredoninae in the Sphecidae, and of the megachilid bees have this connection opened early in the larval life and excrete small meconial pellets during the last three instars.

Many larvae of the Parasitica and those of the social parasites among the Aculeata are hypermetamorphic, that is, the first instar larva is very different in appearance and behavior from the succeeding instars, which assume the typically maggotlike form of mature larvae of the Apocrita. There are at least 10 different primary larval forms among the species of Apocrita exhibiting hypermetamorphosis. Rarely, the second instar larva may be different in appearance from either the first or third.

There are few reliable data on the number of larval instars. Rather meager information indicates that there are three or four instars in various groups of Aculeata.

Silken cocoons are of general occurrence throughout the order, although they are absent in most Chalcidoidea and Cynipoidea, some Formicoidea, and absent sporadically throughout the rest of the Aculeata. The cocoons may be of two layers, as in a few sawflies and wasps, but normally there is only one layer of silk or the cocoon may consist of only a silken cap. Foreign material may be frequently incorporated in the cocoon. The larva may impregnate the silk with secretions from the gut which cause the cocoon wall to become brittle and varnished. Sand, mud, or prey remnants may be incorporated in the silk of the cocoon. Almost all Hymenoptera that spin cocoons produce the silk from labial glands through the mouthparts, but those Chalcidoidea that construct cocoons produce the silk in the Malpighian tubules and spin it from the anus.

Economically important species. Many of the Hymenoptera are of economic importance and include both useful and destructive forms. The species listed in the following table are North American except where noted to the contrary. Common names are given, and a brief note is included explaining the importance of each species.

Table 2. Economically important Hymenoptera

Family	Scientific name	Common name	Economic importance
Pamphiliidae	*Neurotoma inconspicua*	Plum, web-spinning sawfly	Larvae spin webs and eat foliage of plums
	Pamphilius persicus	Peach sawfly	Larvae roll peach leaves and feed on them
Cimbicidae	*Cimbex americana*	Elm sawfly	Larvae eat foliage of elm, willow, poplar, and maple
Diprionidae	*Diprion hercyniae*	European spruce sawfly	Larvae defoliate spruce; introduced from Europe
	Neodiprion lecontei	Red-headed pine sawfly	Larvae defoliate pines
Tenthredinidae	*Caliroa cerasi*	Pear slug	Larvae skeletonize foliage of pear, cherry, plum; probably introduced from Europe
	Fenusa ulmi	Elm leaf miner	Larvae mine leaves of elm; probably introduced from Europe
	Cladius isomerus	Bristly rose slug	Larvae skeletonize rose leaves
	Hoplocampa cookei	Cherry fruit sawfly	Larvae feed inside cherries
	Pteronidea ribesii	Imported currant worm	Larvae feed on leaves of currants, gooseberries; introduced from Europe
Siricidae	*Tremex columba*	Pigeon tremex	Larvae bore in trunks of weakened or dead maple, oak, elm, other deciduous trees
Cephidae	*Hartigia trimaculata*		Larvae bore in stems of roses, blackberries
	Cephus cinctus	Wheat stem sawfly	Larvae bore in stems of wheat, rye, timothy, and wild grasses
Braconidae	*Aphidius* spp		Larvae are internal parasites of aphids
	Meteorus spp.		Larvae are internal parasites of lepidopterous and coleopterous larvae
	Macrocentrus spp.		Larvae are internal parasites of lepidopterous larvae
	Apanteles spp.		Larvae are internal parasites of lepidopterous larvae
	Opius spp.		Larvae are dipterous parasites
	Spathius spp.		Larvae are external parasites of coleopterous larvae
Ichneumonidae	*Scambus* spp.		Larvae are internal parasites of small lepidopterous larvae in leaf mines, leaf rolls, galls
	Polysphincta spp.		Larvae are external parasites of spiders

Table 2. Economically important Hymenoptera (Cont.)

Family	Scientific name	Common name	Economic importance
Ichneumonidae (Cont.)	*Ephialtes* spp.		Larvae are internal parasites of lepidopterous pupae
	Megarhyssa macrurus		Larvae are external parasites of wood-boring sawfly larvae
	Gelis spp.		Larvae are parasites in cocoons of other Ichneumonoidea and in egg sacs of spiders
	Acroricnus aequalus		Larvae are parasites in nests of mud-dauber wasps
	Diplazon laetatorius		Eggs are laid in syrphid (Diptera) eggs or young larvae, and adults emerge from host puparia
Mymaridae	*Anagrus armatus*		Larvae are parasites in eggs of Hemiptera
Trichogrammatidae	*Trichogramma minutum*		Larvae are parasites in eggs of other insects
Eulophidae	*Sympiesis* spp.		Larvae are parasites of leaf-mining coleopterous and lepidopterous larvae
	Tetrastichus spp.		Larvae are parasites of a wide variety of insect eggs and larvae including other Chalcidoidea
Thysanidae	*Thysanus* spp.		Larvae are parasites of Homoptera or of other Chalcidoidea which parasitize Homoptera
Encyrtidae	*Copidosome gelechiae*		Polyembryonic parasites of gall-making lepidopterous larvae
	Aphycus spp.		Larvae are parasites of scale insects
	Ooencyrtus spp.		Larvae are parasites in eggs of other insects
Agaontidae	*Blastophagus psenes*	Fig wasp	Lives within figs and fertilizes them; introduced from Europe
Torymidae	*Megastigmus* spp.		Larvae live in seeds
Eurytomidae	*Harmolita tritici*	Wheat jointworm	Larvae bore in stems of wheat
	Harmolita grandis	Wheat strawworm	Larvae bore in stems of wheat
	Bruchophagus gibbus	Clover seed chalcid	Larvae develop in seeds of clover, alfalfa; possibly an introduced species
Cynipidae	*Diplolepis rosae*	Mossy rose gall	Larvae develop in galls on rose stems, introduced from Europe
	Acraspis erinacei	Oak hedgehog gall	Agamic generation (no males), develops in hedgehog gall on oak leaves; sexual generation develops in soft galls in oak buds
	Amphibolips confluenta	Large oak-apple gall	Larvae develop in large globular galls on oak leaves
Evaniidae	*Evania appendigaster*		Parasitic in egg cases of cockroaches; introduced
Scelionidae	*Telenomus* spp.		Parasitic in eggs of many orders of insects
Chrysididae	*Chrysis* spp.	Cuckoo wasps	Social parasites of other wasps and bees
Bethylidae	*Cephalonomia tarsalis*		Parasites of beetle larvae infesting stored grains
	Scleroderma spp		Parasites of old-house borer (Coleoptera); have very painful sting
	Goniozus spp.		Parasites of lepidopterous leaf rollers
Tiphiidae	*Tiphia vernalis*	Spring tiphia	Parasites of Japanese beetle larvae; introduced from Japan, Korea, China
Mutillidae	*Dasymutilla occidentalis*	Cow killer	Parasite of bumblebee pupae
Formicidae	*Eciton* spp.	Legionary ants	Predaceous on other insects
	Pheidole spp	Harvesting ants	Nest in ground and feed on seeds
	Monomorium pharaonis	Pharaoh ant	A pest ant which nests in buildings, introduced
	Solenopsis saevissima var. *richteri*	Imported fire ant	Nests in soil; very aggressive; introduced from South America
	Tetramorium caespitum	Pavement ant	Infests houses; introduced
	Atta spp.	Fungus ants	Nest in soil and feed upon fungi which they cultivate on beds of masticated leaves
	Tapinoma sessile	Odorous house ant	Infests houses
	Camponotus spp.	Carpenter ants	Many species nest in wood

Table 2. Economically important Hymenoptera (Cont.)

Family	Scientific name	Common name	Economic importance
Formicidae (*Cont.*)	*Lasius interjectus*	Larger yellow ant	Nests around building foundations and beneath cellar floors
	Myrmecocystus spp.	Honey ants	Nest in soil in Southwest; some workers store large quantities of honey until abdomen is greatly distended
	Formica exsectoides	Allegheny mound ant	Nests in ground beneath large mounds of excavated soil
	Polyergus spp	Slave-making ants	Enslave workers of other ants to build their nests and feed their young
Vespidae	*Vespa crabro germana*	Giant hornet	Social wasp nesting in sidings of buildings and hollow trees; introduced from Europe
	Vespula (Vespula) spp.	Yellow jacket	Social wasps nesting in ground
	Vespula (Dolichovespula) maculata	Bald-faced hornet	Social wasp which builds large paper nests in trees and shrubs
	Polistes spp.	Paper wasps	Social wasps which build a single umbrella-shaped comb, frequently under eaves or porch roofs
	Eumenes spp.	Jug wasp	Solitary wasps which build small clay jugs in which to rear their young; prey on caterpillars
Pompilidae	*Pepsis* spp	Tarantula hawks	Prey on tarantulas
Sphecidae	*Trypoxylon politum*		Builds the familiar clay "pipe-organ" nests in which it stores many small spiders
	Chlorion ichneumoneum	Great golden digger wasp	Nests in soil and preys on long-horned grasshoppers
	Sceliphron caementarium	Black and yellow mud dauber	Builds clay cells in which it stores many small spiders
	Sphecius speciosus	Cicada killer	Nests in soil and preys on adult cicadas
	Stictia carolina	Horse guard	Nests in ground and preys on a variety of flies
Colletidae	*Colletes* spp.		Solitary bees which nest in ground, frequently gregariously, and collect pollen and nectar from a variety of flowers
Andrenidae	*Andrena* spp.		As in Colletidae
Halictidae	*Halictus* spp		As in Colletidae, except that some species are subsocial or social
Megachilidae	*Megachile* spp.	Leaf-cutter bees	Solitary bees which nest in cavities in wood or in ground, and make cells from pieces cut from leaves
Apidae	*Anthophora* spp.	Mining bees	
	Xylocopa virginica	Carpenter bee	in
	Bombus spp.	Bumblebees	Social bees which nest on or in ground, frequently in old mouse nests
	Trigona spp.	Stingless honeybees	The honeybee of the Tropics; does not occur in North America
	Apis mellifera	Honeybee	The beekeeper's bee; escaped bees nest in hollow trees or house sidings; introduced from Europe

PHYLOGENY

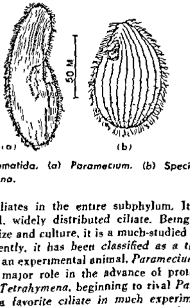
The earliest known fossil Hymenoptera are from the Middle Jurassic. Only Symphyta and Parasitica are known from these strata. The absence of Aculeata perhaps substantiates the claim that this section evolved from the Parasitica; or its absence may be only apparent and due to the meager fossil record. Unfortunately, the wing venation of these fossil sawflies is quite specialized and offers no clues to the derivation of the order. The consensus is that the Hymenoptera probably arose from an ancestral type that also gave rise to the Neuroptera and other related orders. The Baltic amber, which dates from the Tertiary, preserved many ants and a few bees and other Hymenoptera. The bees are all members of genera which are now extinct, but some of the ants are thought to be the same as species living now. [K.V.K.]

Bibliography: H. Bischoff, *Biologie der Hymenopteren*, 1927; C. P. Clausen, *Entomophagous Insects*, 1940; J. H. Comstock, *An Introduction to Entomology*, 9th ed., 1940; K. von Frisch, *The Dancing Bees*, 1955; A. D. Imms, *A General Textbook*

of Entomology, 9th ed., 1957; C. F. W. Muesebeck, K. V. Krombein, H. K. Townes, et al., *Hymenoptera of America North of Mexico; Synoptic Catalog*, USDA Agr. Monograph 2, 1951; C. R. Ribbands, *The Behaviour and Social Life of Honeybees*, 1953; O. W. Richards, *Handbooks for the Identification of British Insects*, vol. 6, pt. 1, 1956; W. M. Wheeler, *Ants; Their Structure, Development and Behavior*, 1926.

Hymenostomatida

An order of the Holotricha which contains many species of Ciliata that often are of small size and fairly uniform ciliation. Primarily, they are of importance as the first possessors of a definite, though inconspicuous, buccal ciliature. This ciliature consists of an undulating membrane (*UM*) on the right side of the buccal cavity and an adoral zone of membranelles (*AZM*) which is primitively composed of three membranelles on the left side. This tetrahymenal, or four-part, buccal ciliary apparatus is considered the fundamental condition from which the oral ciliature of many subsequent higher groups evolved (see CILIOPHORA). The majority of hymenostomes are free-living, fresh-water forms. The parasitic genus *Ichthophthirius* is often mistakenly considered to be a gymnostome (see GYMNOSTOMATIDA). *Paramecium* is the best-known genus of ciliates in the entire subphylum. It is a good-sized, widely distributed ciliate. Being easy to recognize and culture, it is a much-studied form. Until recently, it has been classified as a trichostome. As an experimental animal, *Paramecium* has played a major role in the advance of protozoan genetics. *Tetrahymena*, beginning to rival *Paramecium* as a favorite ciliate in much experimental work, owes its scientific popularity primarily to its ability to grow axenically, that is, free from all other organisms, in a chemically defined medium. Its species are the first animal organisms, excluding the green plantlike flagellated protozoans, to be so grown. This represents a great forward step in experimental studies, particularly those of a biochemical nature, in cancer research and other important fields of direct, immediate interest to man. See AXENIC CULTURE; HOLOTRICHA. [J.O.C.]



Hymenostomatida. (a) Paramecium. (b) Species of Tetrahymena.

Hyperbola

A curve cut from a cone of revolution by a plane that intersects both nappes of the cone and does not contain the apex (see CONIC SECTION). In analytic geometry it is shown that a hyperbola is the locus of points P in a plane, such that $PF = \epsilon \cdot PD$, where PF and PD denote the distances of P from a fixed point F (focus) and a fixed line (directrix) of the plane, respectively, and ϵ is a constant, greater than 1. It is also the locus of points P, the difference of whose distances from two fixed points F, F' (foci) $PF - PF'$ is a constant $2a$ that is less than the distance $2c$ between the foci. The curve is symmetric to the line $g(F,F')$ determined by F, F' and to O, their midpoint. It consists of two branches that are images of each other in the line g through O, perpendicular to $g(F,F')$. There are two lines through O, making equal angles with $g(F,F')$, and to each of which points on each branch get indefinitely close; that is, if point P traverses either branch of the hyperbola its distance from these lines approaches zero. The lines are called asymptotes of the hyperbola. If the asymptotes are mutually perpendicular, the hyperbola is called rectangular or equilateral, since then its transverse axis $AA' = 2a$, where A,A' are on $g(F,F')$ with $OA = OA'$, equals its conjugate axis $BB' = 2b = 2(c^2 - a^2)^{1/2}$, ($B,B'$ on g, $OB = OB'$) The eccen-

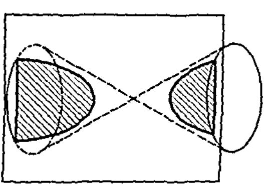


Fig. 1 Hyperbola as a conic section.

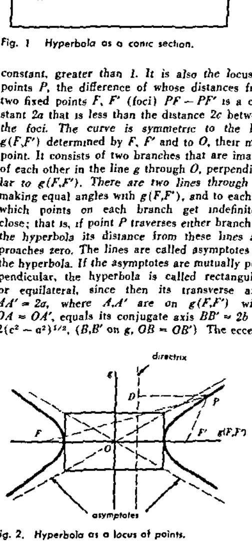


Fig. 2. Hyperbola as a locus of points.

tricity $\epsilon = c/a$ of every equilateral hyperbola is $\sqrt{2}$. The asymptotes of each hyperbola are the diagonals of a rectangle with center at O, having two sides of length $2a$ parallel to $g(F,F')$, and the other two parallel sides have length $2b$. If an ellipse and a hyperbola have the same foci F, F', they are called confocal. Through each point of the plane there is exactly one ellipse and one hyperbola with the same given foci. They intersect each other at right angles at each of the four points they have in common.

The area of the triangle formed with the asymptotes of a hyperbola by a variable tangent is equal to ab. Conversely, if a line cuts the asymptotes of a hyperbola in two points that lie on the same side of the conjugate axis and forms with them a triangle of area ab, then the line is tangent to the hyperbola. The tangent and normal at a point P bisect the angles formed by the lines joining P to the foci F, F'. See ANALYTIC GEOMETRY; HYPERBOLOID. [L.M.BL.]

Hyperbolic function

The hyperbolic sine and cosine of a real or complex variable z are defined by

$$\sinh z = \frac{e^z - e^{-z}}{2} \qquad \cosh z = \frac{e^z + e^{-z}}{2} \qquad (1)$$

Both sinh z and cosh z have the period $2\pi i$ of e^z. From $De^z = de^z/dz = e^z$,

$$D \sinh z = \cosh z \qquad D \cosh z = \sinh z$$

Since $e^z = \sum_{n=0}^{\infty} z^n/n!$ by definition,

$$\sinh z = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!} \qquad \cosh z = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!}$$

the series converge for all z and yield the relations

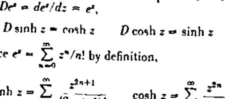
$$\sinh iz = i \sin z \qquad \cosh iz = \cos z$$
$$\sin iz = i \sinh z \qquad \cos iz = \cosh z$$

Thus relations between the circular functions become hyperbolic (and vice versa) when z is replaced by iz. In particular $\cos^2 z + \sin^2 z = 1$ becomes $\cosh^2 z - \sinh^2 z = 1$; and the addition theorems

$$\sin (z + \zeta) = \sin z \cos \zeta + \cos z \sin \zeta$$
$$\cos (z + \zeta) = \cos z \cos \zeta - \sin z \sin \zeta$$

become

$$\sinh (z + \zeta) = \sinh z \cosh \zeta + \cosh z \sinh \zeta$$
$$\cosh (z + \zeta) = \cosh z \cosh \zeta + \sinh z \sinh \zeta$$



Circle, $x^2 + y^2 = 1$; hyperbola, $x^2 - y^2 = 1$.

Moreover Eqs. (1) yield the Eulerian equations

$$i \sin z = \frac{e^{iz} - e^{-iz}}{2} \qquad \cos z = \frac{e^{iz} + e^{iz}}{2} \qquad (2)$$

With $z = x + iy$, the addition theorems give

$$\sinh z = \sinh x \cos y + i \cosh x \sin y$$
$$\cosh z = \cosh x \cos y + i \sinh x \sin y$$

Hence
$$|\sinh z|^2 = \sinh^2 x + \sin^2 y$$
$$|\cosh z|^2 = \sinh^2 x + \cos^2 y$$

and all zeros of sinh z and cosh z lie on the axis $x = 0$ at the points $y = 2n\pi i$ and $y = (2n + \frac{1}{2})\pi i$ respectively. All zeros are simple.

The functions e^z, sin z, cos z, sinh z, cosh z are analytic throughout the complex plane; all have an essential singularity at $z = \infty$; e^z has no zeros, sin z and cos z vanish only on the axis of reals, sinh z and cosh z only on the axis of imaginaries.

The hyperbolic tangent and cotangent are defined by

$$\tanh z = \frac{\sinh z}{\cosh z} \qquad \coth z = \frac{\cosh z}{\sinh z}$$

They have the period πi and have simple poles at the zeros of cosh z and sinh z respectively. Moreover

$$D \tanh z = \frac{1}{\cosh^2 z} \qquad D \coth z = \frac{-1}{\sinh^2 z}$$

By definition

$$\operatorname{sech} z = \frac{1}{\cosh z} \qquad \operatorname{csch} z = \frac{1}{\sinh z}$$

The hyperbolic functions are related to the equilateral hyperbola in much the same way that the circular functions are related to the circle (equilateral ellipse). For both the hyperbola

$$x = \cosh t \qquad y = \sinh t$$

and circle

$$x = \cos t \qquad y = \sin t$$

the area of the sector OAP is the circuit integral

$$\sigma = \tfrac{1}{2} \oint_{OAP} (x\,dy - y\,dx) = \tfrac{1}{2} \int_0^t dt = \tfrac{1}{2} t$$

In both cases the parameter $t = 2\sigma$. See HYPERBOLA; TRIGONOMETRY, PLANE. [L.BR.]

Hyperbolic navigation system

A navigation system that produces hyperbolic lines of position through the measurement of the difference in times of transmission of radio signals from two or more synchronized transmitters at fixed points. When synchronized signals are received from two transmitting stations, the difference in the times of arrival is constant on a hyperbola having the two transmitting stations as foci (see Fig. 1). The measured time difference locates the receiver on the hyperbolic line of position for that time difference.

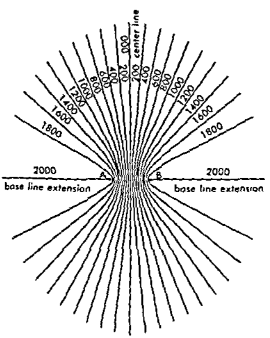


Fig. 1. Lines of constant time difference derived from a pair of transmitters A and B pulsed simultaneously in synchronism.

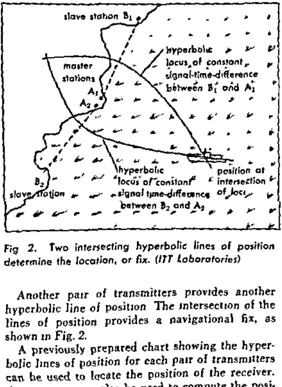


Fig 2. Two intersecting hyperbolic lines of position determine the location, or fix. (ITT Laboratories)

Another pair of transmitters provides another hyperbolic line of position The intersection of the lines of position provides a navigational fix, as shown in Fig. 2.

A previously prepared chart showing the hyperbolic lines of position for each pair of transmitters can be used to locate the position of the receiver. A computer may also be used to compute the position automatically from the received information.

The choice of frequency and locations of the pairs of transmitters determines both utility and accuracy. Base lines must be long to enhance accuracy. Base-line length is also contingent on transmitter power, since sufficient signal strength must be available to maintain accurate synchronism.

The fix is most accurate when the lines of position intersect at right angles. When possible, pairs of stations are so oriented that the lines cross at acute angles in areas of least use. See DECCA; LORAN; OMEGA. [A.A.M.]

Bibliography: J. A. Pierce, A. A. McKenzie, and R. H. Woodward, *Loran*, 1948; P. C. Sandretto, *Electronic Avigation Engineering*, 1958.

Hyperboloid

A quadric surface (that is, a surface having an equation of the second degree in three variables) that has both elliptic and hyperbolic plane sections. It is symmetrical in each of a set of three mutually perpendicular planes that intersect pairwise in lines called axes of the hyperboloid. Two of these planes of symmetry intersect the hyperboloid in hyperbolas. If the third plane of symmetry does not intersect the hyperboloid, the latter consists of two separate but congruent parts and is called an hyperboloid of two sheets. If the third plane of symmetry intersects the hyperboloid, the curve of section is an ellipse (or circle) whose axes are also axes of the hyperboloid, and the latter is a connected surface called an hyperboloid of one sheet. If the axes of an hyperboloid are chosen as coordinate axes and the third plane of symmetry is taken as the xy plane, the equation of the surface may be written in one of the two forms:

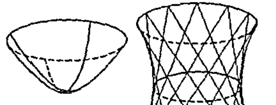
Hyperboloid of one sheet

$$(x/a)^2 + (y/b)^2 - (z/c)^2 = 1$$

Hyperboloid of two sheets

$$(x/a)^2 + (y/b)^2 - (z/c)^2 = -1$$

The numbers a, b, and c are called semiaxes of the hyperboloid, and if $a = b$, the hyperboloid is a hyperboloid of revolution, obtained by revolving an hyperbola about its major axis (two sheets) or minor axis (one sheet). A one-sheeted hyperboloid of revolution can also be generated by revolving one of two skew lines about the other, when both are



Hyperboloids of one and two sheets.

rigidly joined by a line segment along their common perpendicular.

A one-sheeted hyperboloid is a ruled surface, meaning that the tangent plane at each point cuts the surface in a pair of straight lines. String models of this surface may be made by stretching strings along the rulings between parallel elliptical sections. See ANALYTIC GEOMETRY; HYPERBOLA; QUADRIC SURFACE; SURFACE AND SOLID OF REVOLUTION. [J.S.F.]

Hyperfine structure

A splitting of spectral lines due to the magnetic moment of the atomic nucleus. When the lines of the heavier elements are examined with spectrographs of the highest resolving power, many of them are found to consist of two or more closely spaced components. If the element has more than one isotope, some of these components may arise from the nuclides of different mass number (see ISOTOPE SHIFT). This effect may be eliminated by observing the spectra of the pure, separated isotopes; however, one may still observe a splitting if the nuclide has an odd mass number. Valuable information about the nuclear angular momentum (spin), magnetic moment, and electric quadrupole moment is derived from the number and spacing of the energy levels determined from the hyperfine structure. Thus if the nuclear spin is I (in units of $h/2\pi$), the maximum number of components into which any level can be split is $2I + 1$. The separations of these component levels, from which the magnetic and quadrupole moments are derived, may be studied by both optical and radio-frequency methods. See NUCLEAR MOMENTS. [F.A.J.]

Hyperiidea

A suborder of amphipod crustaceans (see AMPHIPODA). Most hyperiids can be recognized by the large eyes which cover nearly the entire surface of the head. Compared to the suborder Gammaridea, the first maxillae and especially the maxillipeds are greatly reduced. In the prehensile pereiopods, the claw is formed by the fifth and sixth segments, the carpus and propodus, rather than by the sixth and seventh segments or the propodus and dactyl, as in the Gammaridea. The second and third somites of the urosome are always fused, a condition rarely found in the Gammaridea.

The Hyperiidea are exclusively pelagic and marine. They are found in all the oceans, from the surface to great depths. Most are characteristic of oceanic rather than neritic waters, although certain species frequent inshore waters in the tropics. After the Copepoda and Euphausiacea, the Hyperiidea are the most abundant planktonic crustaceans.

Little is known about the habits of the hyperiids. Some species are frequently found in association with other animals. Members of the genus *Vibilia* are often associated with salps, and may enter the body of the salp and move about freely. Some species of *Hyperia* and *Hyperoche* are commonly found on Scyphomedusae, clinging to the surface of the bell. The young of *Hyperia galba* have been found in the subgenital pits. *Eupronoe* also sometimes occurs with medusae.

The adult females of *Phronima sedentaria* occur in deep water in gelatinous barrel-shaped cases open at both ends. These cases are said to be fashioned from the tests of salps. The young hatch in the cases and pass through several molts before leaving them to move nearer to the surface.

Other species are truly free-living. Except in tropical and subtropical seas, species of *Parathemisto*, formerly *Themisto*, are the most abundant hyperiids. *P. libellula* occurs in large numbers in Arctic waters and is an important food for ringed seals. Japanese sardines feed on *P. japonica*. *Phrosina semilunata*, *Phronima sedentaria*, *Streetsia challengeri*, and *Brachyscelus crusculum* are eaten by albacore.

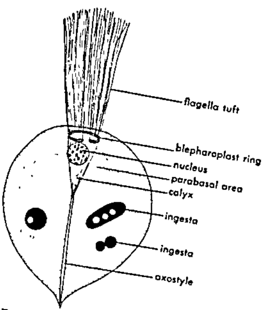
[illegible]

LUMINESCENCE).

The suborder Hyperiidea is divided into two superfamilies, Physosomata and Genuina. Members of the Physosomata, except for a number of species of *Scina* and one species of *Lanceola*, are bathypelagic. The eyes are small or, rarely, absent, and the inner plates of the maxillipeds are free at the apex. The Genuina include the more familiar hyperiids with large eyes and completely fused maxillipedal inner plates, and are divided into three groups according to the structure of the male antennae. In the Recticornia, containing the genera *Vibilia*, *Paraphronima*, and *Cystisoma*, the first antennae are straight. They arise from the anterior margin of the head, and have few flagellar segments. The first antennae of the Filicornia, in such genera as *Hyperia*, *Dairella*, *Primno*, and *Phronima*, are also inserted anteriorly, and have many-segmented flagella. The antennae of the Curvicornia originate from the ventral margin of the head. This may be observed in *Eupronoe*, *Lycaea*, *Oxycephalus*, and *Platyscelus*, in which the first segment is large and curved, with the remaining segments few in number. The flagellar segments of the second antennae are folded on themselves. The Curvicornia comprise globular forms with greatly enlarged basal segments on the fifth and sixth legs as well as slender, elongate species, culminating in the rod-shaped *Rhabdosoma*. [T.E.B.]

Hypermastigida

The most complex flagellates, both structurally and in modes of division. All inhabit the alimentary canal of termites, cockroaches, and woodroaches. These organisms are multiflagellate, often with complicated blepharoplast-parabasalaxostylar structures. The nucleus is single and the organisms are plastic and slow-moving, generally ovoid to elongate. Flagella occur in spiral rows, in tufts, or over the entire body. These flagellates vary from 15



Lophomonas blattarum.

to 350 μ in size. They may be holozoic or saprozoic. Ingestion is usually pseudopodial, in the posterior region. A high degree of adaptiveness exists and species peculiar to one host species are not viable in another. Sexual processes and multiple fission are known, but trinary fission is most common. *Holomastigotoides* has spirochetes attached to its trophic body zone.

Lophomonas inhabits the cockroach intestine and is apparently cosmopolitan. The body is round to pyriform with a tuft of apical flagella arising from a blepharoplast ring which forms a calyx enclosing a nucleus. Six to eight chromosomes form on a spindle of which the poles are centrioles. Laboratory culture is easy. [J.B.L.]

Hyperon

Any elementary (noncomposite) particle intermediate in mass between the neutron and the deuteron. The known hyperons, listed in the accompanying table, appear grouped as charge multiplets whose members have comparable mass values but different electric charges. This structure suggested to M. Gell-Mann that the strong interactions of the hyperons must be charge-independent, each multiplet being characterized by an isotopic spin T such that the multiplet has $2T + 1$ charge states. All of the evidence on hyperon reactions is in accord with this hypothesis. *See* ELEMENTARY PARTICLE; ISOTOPIC SPIN; PARITY (QUANTUM MECHANICS).

Known types. The Λ-hyperon is a neutral particle of mass $2183m_e$ (m_e is the electron mass), the first of the so-called V-particles to be established as a distinct species (*see* V-PARTICLE). It is the only member of the Λ-multiplet, so that isotopic spin $T = 0$ holds for the Λ-hyperon. It decays to a pion and a nucleon with a mean lifetime of 2.5×10^{-10} sec. The observation of forward-backward asymmetry in the directional distribution of the outgoing pions relative to the spin direction for the parent Λ-hyperon has established a strong violation of the principle of parity conservation for the decay mode $\Lambda \rightarrow \pi^- + p$. The existence of attractive forces of nuclear strength between the Λ-hyperon and nucleons is established by the existence of Λ-hypernuclei or Λ-hyperfragments (consisting of a Λ-hyperon bound to an ordinary nucleus), which remain stable until the decay of the Λ-hyperon.

The Σ-hyperons form a triplet (isotopic spin $T = 1$), the Σ^--, Σ^0-, and Σ^+-hyperons of mean mass $2334m_e$. The Σ^+- and Σ^--hyperons have lifetimes about 10^{-10} sec, their decay leading to a pion and a nucleon; the Σ^0-hyperon decays very rapidly (lifetime $\sim 10^{-19}$ sec), transforming into a Λ-hyperon with emission of a photon. The Σ-hyperons also interact strongly with nucleons, as shown, for example, by the observation of Σ^--proton elastic scattering through large angles. A Σ-hyperon reacts rapidly with a nucleon of suitable charge, transforming into a Λ-hyperon with release of about 80 Mev kinetic energy; for example,

$$\Sigma^- + p \rightarrow \Lambda + n \qquad \Sigma^+ + n \rightarrow \Lambda + p$$

are known reactions. These absorptive reactions preclude the existence of Σ-hypernuclei except for several possibilities, such as $p\Sigma^+$, $pp\Sigma^+$ or $n\Sigma^-$, $nn\Sigma^-$, it is not yet settled whether these latter systems form bound states.

The Λ- and Σ-hyperons are formed in closely related interactions; for example, their formation

The known hyperons and their properties

Hyperon	Λ	Σ^-	Σ^0	Σ^+	Ξ^-	Ξ^0
Mass, m_e	2183 0 ± 0 2	2341.4 ± 0.7	2332.7 ± 1.0	2328 0 ± 0.4	2581 5 ± 1.0	2565 ± 15
Spin	½	½	½?	½	?	?
Isotopic spin	0	1	1	1	½?	½?
Strangeness number	−1	−1	−1	−1	−2	−2
Lifetime, 10^{-10} sec	2.50 ± 0.09	1.60 ± 0.10	$\sim 10^{-19}$ sec	0.75 ± 0.10	2.0 ± 0 6	?
Decay modes and their relative frequency	$\pi^- + p$, 65% $\pi^0 + n$, 35%	$\pi^- + n$ 100%	$\Lambda + \gamma$ 100%	$\pi^0 + p$, 50% $\pi^+ + n$, 50%	$\pi^- + \Lambda$ 100%	$\pi^0 + \Lambda$?

in pion-proton collisions of sufficiently high energy or in K^- absorption by protons,

$$\pi^- + p \rightarrow \begin{cases} K^0 + \Lambda \\ K^0 + \Sigma^0 \\ K^+ + \Sigma^- \end{cases} \qquad K^- + p \rightarrow \begin{cases} \pi^- + \Sigma^+ \\ \pi^0 + \Sigma^0 \\ \pi^+ + \Sigma^- \\ \pi^0 + \Lambda \end{cases}$$

has been especially well studied. These reactions follow the law of conservation of strangeness, provided a strangeness quantum number $s = -1$ is assigned both to Λ- and to Σ-hyperons (*see* MESON). This suggests that the Λ- and Σ-multiplets may have a deep relationship, as their relatively small mass difference also suggests. Both multiplets consist of fermions of intrinsic spin $\frac{1}{2}\hbar$ ($\hbar$ is Planck's constant divided by 2π). It is conventional (but holds no physical consequences) to assign positive parity to the Λ-hyperon; the parity of the Σ-hyperons relative to this has not yet (early 1960) been established.

The Ξ-hyperons (mass about $2580m_e$) form a doublet (isotopic spin $T = \frac{1}{2}$), consisting of Ξ^-- and Ξ^0-hyperons. About 50 examples of the Ξ^--hyperon (also known as the cascade particle) had been reported up to 1960. It decays to $\Lambda + \pi^-$ with mean lifetime about 2×10^{-10} sec. The first clear example of Ξ^0-hyperon production, through the reaction

$$K^- + p \rightarrow \Xi^0 + K^0$$

was reported in 1959; the Ξ^0-hyperon lived for 1.5×10^{-10} sec and decayed to $\Lambda + \pi^0$. Ξ^- production has been observed in both K^--nucleon and pion-nucleon collisions. The latter process involves the simultaneous formation of two K-mesons; for example,

$$\pi^- + p \rightarrow \Xi^- + K^+ + K^0$$

Strangeness conservation holds for these reactions provided $s = -2$ is assigned to the Ξ-hyperon doublet. Neither the Ξ spin nor the Ξ parity is yet known.

Interaction with matter. Negative hyperons coming to rest in matter first undergo capture into atomic orbits, then cascade down to inner orbits from which they are absorbed by the nucleus through reactions which release energy and cause breakup of the *nucleus*, giving rise to nuclear stars. The Σ^--capture stars are the result of the absorptive reaction $\Sigma^- + p \rightarrow \Lambda + n$. The majority ($\sim 75\%$) of Σ^--stars do not emit charged particles; the stars visible in photographic emulsion are generally small (visible energy 10–20 Mev), and frequently (about 10% of Σ^--*captures*) emit a Λ-hyperfragment.

It is expected that Ξ^--capture stars should result from the absorptive reaction

$$\Xi^- + p \rightarrow \Lambda + \Lambda$$

This prediction follows from the conservation laws well tested for other K-meson and hyperon reactions. No example of a Ξ^--star has yet been reported.

Heavier hyperons. Several cosmic ray events have been reported which suggest that still heavier hyperons may exist, but these are at present isolated events whose interpretation is not completely unambiguous. No evidence suggesting the existence of heavier hyperons has yet been obtained from accelerator experiments. [R.H.D.]

Bibliography: M. Gell-Mann and E. P. Rosenbaum, Elementary particles, *Sci. American*, 197: 72–88, 1957; M. Gell-Mann and A. H. Rosenfeld, Hyperons and heavy mesons, *Ann. Rev. Nuclear Sci.*, 7:407–478, 1957; J. G. Wilson (ed.), *Progress in Cosmic Ray Physics*, vol. 3, 1956.

Hyperplasia

An increase in the number of cells. It should be distinguished from hypertrophy which is an increase in the size of cells, although both bring about an increase in the size of an organ. The term hypertrophic is used with reference to the enlarged organ. Organs, however, do not undergo hyperplasia; the term is necessarily restricted to cells. The average size of cells is much the same throughout the animal kingdom despite the vast difference in the size of individuals of different species so that hyperplasia rather than hypertrophy distinguishes the elephant from the mouse.

The ability to divide or undergo mitosis is limited to certain types of cells known as intermitotic cells. The term intermitotic means that they lead an intermitotic existence; they arise from division of a parent cell and terminate their lives in mitosis to produce two daughter cells. Cells that are incapable of division are known as postmitotic cells; they lead a postmitotic existence and their lives are terminated by death. Generally speaking, fully functional cells are postmitotic, while partially differentiated, reserve, or stem cells are intermitotic. There is much evidence to suggest that, in response to exactly the same stimulus, intermitotic cells undergo hyperplasia, and postmitotic cells hypertrophy.

Certain types of hyperplasia may be extremely difficult to differentiate from neoplasia, both grossly and microscopically. Not infrequently the mass of hyperplastic cells causes a localized swelling which grossly resembles a tumor. As in neoplasia, a rapid rate of proliferation is reflected in the large number of mitoses which may be evident. Hyperplasia differs from malignant neoplasia, however, in that the cells appear to be differentiating in an orderly fashion and the features of anaplasia are lacking. *See* ANAPLASIA; ATROPHY; NEOPLASIA. [W.F.P.]

Hypersensitivity

An abnormal reactivity in man and animals to allergens (substances foreign to the body) induced by exposure through injection, inhalation, ingestion, or contact. It is usually referred to as allergy (altered reactivity), and its commonly known clinical forms include asthma, hay fever, hives,

infantile eczema, and the contact skin lesions caused by poison oak and poison ivy. Since the allergen may itself be nontoxic, the first experience with it occasions no difficulties. Second and subsequent exposures, however, may give rise to a variety of reactions of different degrees of severity, ranging from transient hives to death. There are a number of instances in which reactivity is not so superficially apparent, but where damage to tissues and organs may be severe. This is true in tuberculosis, and of the lesions of rheumatic fever, glomerulonephritis, disseminated lupus erythematosus, and several other serious systemic disorders.

The allergen, or antigen, induces the allergic state, and subsequently becomes responsible for the occurrence of reactions. Allergens vary in chemical nature, ranging from proteins and polysaccharides through simpler drugs and chemical compounds, even the salts of metals. The simpler substances (haptens) cannot function as allergens in inducing the hypersensitive state unless they become linked to proteinaceous substances of the tissues. But the simple substances themselves may be sufficient for triggering the hypersensitive reaction in the already sensitized subject.

Hypersensitivity may be classified into two categories, delayed and immediate. The delay has nothing to do with the induction of the allergic state; this period is about the same for all immunologic responses. It refers to the fact that the reactions begin after several hours, rather than in seconds or minutes as in immediate hypersensitivity. There are two chief distinctions between these two kinds of hypersensitivity. (1) In delayed hypersensitivity the tissues involved in the reaction may be of any kind and, accordingly, delayed reactions may damage any kind of cells. (2) Since the tissue cells themselves react without any connection to the circulating antibodies, the delayed hypersensitive state cannot be transferred from the allergic subject to a normal one by injections of serum, as it can in immediate hypersensitivity

Delayed hypersensitivity. The two kinds of delayed hypersensitivity are infection hypersensitivity and noninfection hypersensitivity. Infection hypersensitivity results from infection with bacteria, viruses, fungi, or animal parasites. Although all infections induce delayed hypersensitivity, this is especially pronounced in some instances. An example is the allergy of tuberculosis; the destruction of [illegible]

and systemic fungal diseases such as coccidiomycosis. Isolated allergens of the infectious agents will not give rise to the hypersensitive state; a complex of substances from the agent is necessary to induce this type of allergy In the case of the tubercle bacillus, the protein allergen of the organism must be accompanied by a nonallergenic lipoidal component of the organism in order to establish this reactive state. The allergen alone, however, can cause reactions once hypersensitivity is established.

Noninfection hypersensitivity, the second category of the delayed type, is exemplified by the contact sensitivities resulting from exposure of the skin to a variety of plant and chemical substances. Examples of these are poison ivy, poison oak, primrose, and chemical compounds such as penicillin. It is a peculiar fact that virtually all instances of contact delayed hypersensitivity are induced by allergens of relatively simple chemical nature rather than by the complex, high molecular weight proteins or polysaccharides. This is difficult to understand in view of the superior antigenicity (ability to induce antibodies) of the complex compounds. However, it may be explained on the still speculative basis that these simple substances combine with constituents of the skin in order to become allergenic, and that in this process of chemical alteration of tissue components some material may be liberated from the tissue. This material may have the same ability as the lipoidal material of the tubercle bacillus described above to cause the development of this special kind of hypersensitivity.

Immediate hypersensitivity. Immediate hypersensitivity contrasts with the delayed hypersensitivity in rapidity of reaction, and depends upon the combination of circulating antibodies with allergen. As a result of this combination, intermediary substances such as histamine are released, and certain tissues are affected, notably blood vessel walls and smooth muscle. Disturbances in these tissues in various areas of the body account for many of the superficially different phenomena which occur in immediate hypersensitive reactions. This sensitive state can be transferred to a normal subject by means of serum. Several types of immediate hypersensitivity are anaphylaxis, Arthus reaction, and atopic allergy.

Passive transfer of immediate hypersensitive states can be accomplished by injecting the serum from a sensitized into a normal subject. After some hours the recipient becomes sensitive to the allergen just as if he had been actively sensitized. In one form of immediate hypersensitivity, the Arthus type, no interval for sensitization is required, and it is thought that this particular allergic reaction depends upon the interaction of antigen and antibody in the blood stream itself. In contrast, in anaphylaxis interaction is presumed to occur at the surface of cells where tissue-bound antibody is present. Delayed hypersensitivity cannot be transferred by means of serum. A suggestion has been made, but not confirmed, that α-globulin of the serum contains antibodies capable of transferring such sensitivity. It has been amply demonstrated that delayed hypersensitivity may be transferred in animals by means of spleen and lymph node cells, and in man by blood leukocytes. Lymphocytes may be of most importance in this transfer. It has not been decided whether they function simply as passive

tissue elements in the recipient, giving it sensitivity so long as the transferred cells survive, or whether they induce in the recipient a state of reactivity which is a continuation of the active state engendered in the cell donor.

Histamine undoubtedly plays a role in causing immediate hypersensitive reactions. The wheal (edema), flare and erythema which constitute the hive, has the features of the triple response described by Sir Thomas Lewis as the characteristic dermal response to the injection of histamine, or to superficial trauma which may release histamine in the tissues. It seems fairly certain that the combination of antigen with antibody in vivo may cause the release of histamine from certain sites in the tissue. The site of histamine's greatest concentration is in the granules of mast or basophilic cells (the reticuloendothelial system) distributed through the body. Histamine is a base which exists loosely bound with heparin, an acidic substance. Various compounds, among them the antigen-antibody complex, may serve as histamine and heparin releasers. Histamine causes dilatation of capillaries and contraction of smooth muscle. It has other pharmacologic activities also, but these two account for the previously described aspects of the hive. The release of heparin is shown by a decreased coagulability of the blood of subjects in anaphylactic shock. Other released intermediary substances include serotonin or hydroxytryptamine, acetylcholine, and a substance known as SRS, or slow reactive substance. These can all be involved in muscle contraction.

A nonallergic reaction resembling anaphylaxis is known as an anaphylactoid reaction. In a general sense, a reaction which mimics any hypersensitive manifestation is called anaphylactoid. Such a reaction depends upon the primary toxicity of the inductant rather than the subject's immunologic responsiveness to it. Certain reactions which are loosely termed hypersensitive may result from drugs which are toxic for some individuals.

The role of steroids in various hypersensitive states is not entirely clear-cut. However, certain interpretations can be made, and practical applications of cortisone and other similar adrenocortical hormones, as well as ACTH, the anterior (pituitary) corticotropic hormone, are routine. Both cortisone and the pituitary hormone are known to have depressive effects upon the formation of antibodies, at least in certain animal species, as well as anti-inflammatory activity. Both these effects are demonstrable in the control of certain hypersensitive states and reactions. Thus, if steroids are given to rabbits during the phase of attempted induction of Arthus reactivity, they may modify or suppress the development of hypersensitivity. On the other hand, if steroids are administered to guinea pigs during the induction of the anaphylactic state, the development is usually not depressed. One may interpret this difference by supposing that in the first case the animal, in order to develop Arthus reactivity, must respond to the antigenic stimulus with a good level of antibody production, and that this can be sufficiently suppressed by the hormones so as to modify reactivity significantly. In the second case, however, the degree of immunologic response required to set up the anaphylactic state is so minor that the activity of the steroids is not sufficient to suppress responsiveness to this point. The steroids are most useful on the basis of their anti-inflammatory effects in alleviating the reactions in already sensitive subjects. They are used clinically in severe atopic hypersensitivities while attempted desensitization is under way, or when the state is intractable to immunologic therapy. Steroids are applicable also in delayed hypersensitivity, when the lesions are so troublesome as to require alleviation before they subside spontaneously. Because of effects of these hormones upon electrolyte balance and carbohydrate metabolism, as well as their potentially dangerous influence upon the course of existing infections, they are not employed promiscuously. *See* ALLERGY, ATOPIC; ANAPHYLAXIS; ARTHUS REACTION; IMMUNOLOGY. [S.R.]

Bibliography: S. Raffel, *Immunity*, 1953.

Hypersonic flight

Flight at speeds well above the velocity of sound. In the span of a few years, aircraft and missile flight speeds have increased tremendously, from far less than the speed of sound (subsonic) to speeds required for orbiting or escape from Earth. By convention, hypersonic flight starts at about Mach 5 (five times the speed of sound) and extends upward in speed indefinitely. A more objective definition, attributed to T. von Karman, states that hypersonic flight starts when the cross-flow Mach number (or Mach number component perpendicular to the longitudinal axis) is 1. In any case, flow about the body is sensibly unchanged as the speed of the body increases further.

Conditions at high speed. Below about 350,000 ft altitude, air is sufficiently dense to be considered a continuum and to form viscous laminar and turbulent boundary layers on body surfaces. In this region, the flight speed divided by the sound speed, that is, the Mach number, is a characteristic aerodynamic parameter. Above about 350,000 ft in the realm of superaerodynamics, air density is so low that mean free path between molecular collisions is large compared to body dimensions, and a new parameter known as the molecular speed ratio (ratio of flight speed to most probable incident molecule speed) is used. *See* SUPERAERODYNAMICS.

Subsonic, transonic, and supersonic vehicles fly at speeds less than, equal to, and greater than the local speed of sound, respectively. However, when Mach number is large, the flow field around the object exhibits a special behavior which is worth studying separately from supersonic flight. It is then called hypersonic flight. For long slender bodies and for thin airfoils at small angles of attack, transition from supersonic to hypersonic flight may require a Mach number equal to or greater than

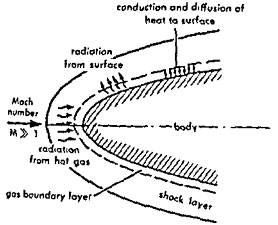


Fig. 1. Hypersonic flow over a blunt-nosed body.

5 (flight speed approximately 5000 ft/sec). On Earth, the fastest-moving natural objects which can be seen by an observer are meteorites traveling at 25,000–125,000 ft/sec, or 0.000256–0.00128 times the speed of light.

A body entering Earth's atmosphere from space (a meteorite, a ballistic missile, or a space glide vehicle, for example) has a high velocity and hence a large quantity of kinetic and potential energy. The kinetic energy equals $WV^2/2g$, and the potential energy equals WZ, where Z is the altitude. At hypersonic speeds, potential energy of the vehicle is small compared to kinetic energy. For instance, the potential energy of a body at an altitude of 200,000 ft about equals the kinetic energy of the body if it is moving at 3600 ft/sec.

During reentry, drag forces act upon the body and cause it to decelerate, thus dissipating the kinetic energy of the body. The portions of kinetic and potential energy which are lost from the vehicle are then transferred to the air within the flow field of the moving object. The flow field around a blunt-nosed object (body of revolution or wing leading edge) generally exhibits a distinct bow shock wave, a shock layer of highly compressed hot gas, and a highly sheared boundary layer over the surface (Fig. 1). The temperature of the shock

electrons begin to be knocked off, the gas becoming ionized. Thus, in general, the working fluid in hypersonic flight cannot be considered to behave as a perfect gas. Heat energy is then transferred from the hot gases to the vehicle by conduction and diffusion of chemical species in the boundary layer, and by radiation from the shock layer near the nose. Heat energy is also radiated from the surface to space. Therefore, one of the important problems which confronts the designer of space reentry vehicles is to build a vehicle which can absorb large heat loads from adjacent hot gas layers, while simultaneously retaining the capability of carrying a useful payload. *See* REENTRY; SPACECRAFT STRUCTURE.

Vehicular trajectories. A hypersonic vehicle in sustained powered flight maintains its speed at a constant level (Fig. 2). It needs only to be accelerated initially to its operating velocity. An engine or rocket motor provides the thrust that overcomes aerodynamic drag, but the propulsive work done to sustain the flight speed must be dissipated as heat to the surrounding air or radiated to space.

There are three types of trajectories for the unpowered portion of the flight of a hypersonic vehicle while reentering Earth's atmosphere.

Ballistic trajectory. A trajectory such as that of an IRBM or an ICBM is actually a segment of one of Kepler's planetary ellipses (Fig. 2). The range of the ballistic missile is shorter than that of other reentry vehicles for given initial conditions; the accompanying table summarizes the range equations of the four principal flight paths.

The short transit time of the ballistic missile in Earth's atmosphere gives rise to a large rate of kinetic energy dissipation. High decelerations are also experienced. The effect of aerodynamic heating on the vehicle can be reduced if a blunt shape

Ratio of range to height for hypersonic flight paths

Boost-glide range (neglecting velocity due to Earth's rotation):

$$\frac{S_{BG}}{R} = \sqrt{\frac{U_i^2 - U_0^2}{1 - U_0^2}} + (L/2D) \ln\left(\frac{1}{1 - U_0^2}\right) \quad (1)$$

Maximum ballistic range (neglecting velocity due to Earth's rotation):

$$\frac{S_{BB_{max}}}{R} = 2 \sin^{-1}\left(\frac{U_i^2}{1 - U_i^2}\right) \quad (2)$$

Boost-skip range (neglecting skipping phase):

$$\frac{S_{BS}}{R} = 2 \sum_{n=0}^{n} \tan^{-1}\left(\frac{\sin\theta_f \cos\theta_f}{e^{4n\theta_f D/L} - U_i^2 \cos^2\theta_f} U_i^2\right) \quad (3)$$

Sustained-propulsion range (ignoring take-off, acceleration, and climb):

$$S = \frac{LV \ln (W_i/W)}{Dc(1 - V^2)} \sqrt{gR} \quad (4)$$

where

c = specific fuel consumption
g = acceleration of gravity
L/D = lift-to-drag ratio
n = number of ballistic phases
R = distance to center of Earth
$S, S_{BB}, S_{BS}, S_{BG}$ = range
$U_i = u_i/\sqrt{gR}$ = dimensionless initial velocity
$U_0 = u_0/\sqrt{gR}$ = dimensionless velocity at peak altitude
$V = v/\sqrt{gR}$ = dimensionless airplane flight speed
$\sqrt{gR}$ = satellite velocity
W_i/W = aircraft and fuel initial-to-final weight ratio
θ_f = reentry angle

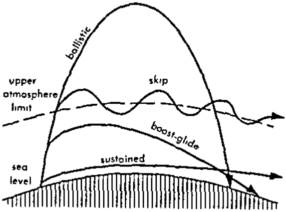


Fig 2 Typical trajectories for hypersonic flight.

is employed. This high drag shape expends a large portion of its kinetic energy in pressure drag rather than in friction forces. The effect of local aerodynamic heating can be further reduced by employing ablation, shock generation, and transpiration cooling techniques.

Skip trajectory. This trajectory is made up of ballistic phases connected by skipping phases (Fig. 2). The principle is that the vehicle develops lift with minimum drag during the skip phases. The flight path resembles a stone skipping over a pond. In the ballistic phase, the vehicle experiences only gravitational and inertial forces; accordingly, each phase is simply a segment of one of Kepler's ellipses. During the skipping phase, large aerodynamic forces are experienced. They tend to be so large when the reentry angle is steep that, for most practical purposes, it is permissible to neglect the gravity force and to treat such a skipping phase as an impact. The heating effect is more severe during the skipping process, even if radiation cooling is employed. A strong structure is required for a skip vehicle because of the heating and violent impact loads during the skip phase for a maximum range trajectory. The skip vehicle with a lift-to-drag ratio between 1 and 4 appears to be more efficient than both the ballistic and glide types in converting high velocity into long range.

Glide trajectory. In a glide trajectory, the vehicle descends and decelerates slowly through Earth's atmosphere in such a manner that the aerodynamic lift plus inertial forces acting on the vehicle just counterbalance the weight and drag in the glide phase. For lift-to-drag ratios greater than about 1.5, the glide trajectory exceeds the maximum ballistic range for the same payload-to-take-off mass ratio, that is, for the same velocity at burnout. For lift-to-drag ratios in the neighborhood of 4 and greater, the glide vehicle is comparable to the skip vehicle in ability to convert velocity into range.

To obtain a relatively high lift-drag ratio, a slender shape is necessary. This low drag configuration will in turn increase the total convective heat transfer to the vehicle because of long flight time, but the effect is not as severe as that encountered by the skip vehicle. The combination of radiation cooling, structural heat sink techniques, and glide trajectory control (gradual descent and deceleration) is able to reduce skin temperatures to a level feasible for present materials. For flight ranges in the order of Earth's radius and greater, the performance of hypersonic vehicles compares favorably with that of the supersonic airplane. *See* BALLISTIC MISSILE; GUIDED MISSILE. [S.Y.C.]

Bibliography: W. D. Hayes, Hypersonic aerodynamics, *Astronautics*, March 24–25, 72, 74, 1959; Hypersonic Aerodynamics Issue, *Jet Propulsion*, vol. 26, no. 4, April, 1956; Hypervelocity Flight Issue, *Jet Propulsion*, vol. 27, no. 11, November, 1957.

Hypertension

The persistent elevation of blood pressures above accepted normal levels. Although there is no absolute agreement on the acceptable range of pressure for an individual, the chronic elevation of diastolic pressure above 90 mm of mercury is usually considered to be abnormal. Systolic pressures above 140 mm of mercury may be abnormal, but in both cases, consideration must be given to age, weight,

sons. However, hypertension is in reality a symptom and not a disease itself, in the strict sense.

Abnormal elevations of blood pressure may occur transiently as the result of increased activity, emotion, or illness. Two basic mechanisms may be present, either singly or together. The first is increased cardiac output; the second and most important is increased peripheral resistance of the blood vessels. *See* CIRCULATORY SYSTEM.

Many common diseases may be accompanied by various degrees of hypertension. These include hyperthyroidism, increased intracranial pressure, arteriosclerosis, and various anxiety states. *See* NEUROSIS; THYROID GLAND DISORDERS.

Although hypertension occurs in many specific diseases, by far the highest incidence falls into the category of benign, essential, or idiopathic hypertension. The exact causes of this most common form of vascular disease are unknown, although contributing and associated factors are often

result of arteriolar spasm or constriction caused by substances circulating in the blood, by neurogenic stimulation, or by degenerative and disease changes in the blood vessels is not clearly understood. Other mechanisms have also been proposed.

The influence of body build, the tendency toward obesity, familial predisposition toward hypertension, and other relationships are known or suspected.

Recent advances in therapy, particularly the development of potent antihypertensive drugs, have given symptomatic relief and reduced the severity of many cases of hypertension. [F.C.ST.]

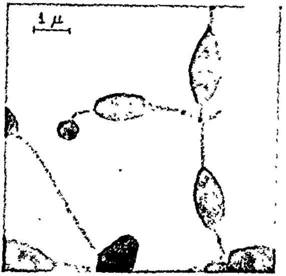
Hypertrophy

The increase in size of a cell, tissue, or organ. Hypertrophy of an organ or tissue is due either to an increase in the mass of the individual cells (true hypertrophy) or to an increase in the number of cells making up the tissue (hyperplasia) or to both. Certain organs such as the heart consist of cells which lack the ability to divide and in such cases hypertrophy results in an increase in the size rather than in the number of the constituent cells. In other tissues such as the liver, lymph nodes, and adrenal cortex, hypertrophy of the organ is effected principally by an increase in the number of cells. See HYPERPLASIA.

The mechanism by which cells or organs hypertrophy is unknown or at least obscure in most instances. When the hypertrophy is the result of physiological rather than pathological stimuli, it is termed physiological hypertrophy. Examples are the increased mass of skeletal muscles which may be produced by physical exercise, the enlargement of the uterus in pregnancy, and the increase in circulating red blood cells found in people living at high altitudes. Compensatory hypertrophy is the increase in size of an organ or organ system following removal or destruction of a portion of it. The most dramatic example is the hypertrophy of one kidney which normally follows removal or destruction by disease of the contralateral kidney. Hypertrophy appears as the consequence of an increased demand for work. In the case of the kidney, it occurs because of the greater filtration and secretory demands which are placed on the remaining nephrons; in the case of the heart in hypertension or valvular disease, because more blood must be pumped per unit time or at a greater pressure. In the case of muscle, an important factor appears to be concerned with stretching, for simple stretching without contraction results in hypertrophy. With smooth muscle, proliferation of new muscle cells, as well as hypertrophy of the existing ones, may also be produced by stretching the tissue. Hypertrophy can also be produced in many organs and tissues by hormonal stimulation. The breast enlarges during pregnancy and lactation as the result of increased levels of pituitary and ovarian hormones. The thyroid gland undergoes hypertrophy and hyperplasia as the result of stimulation by pituitary hormones. The parathyroid glands are frequently seen to hypertrophy because of pituitary stimulation secondary to abnormal chemical changes in the blood. Since both hyperplasia and hypertrophy involve a net increase in cell mass and consequently cell protein, adequate cellular nutrition, including an adequate blood supply, is a prerequisite for such changes. See KIDNEY. [W.F.P.]

Hyphomicrobiales

are found in mud and water in fresh-water ponds—water.



Electron micrograph of Rhodomicrobium vannieli

and may also parasitize *Daphnia*, a fresh-water crustacean. The order contains two families. See SCHIZOMYCETES.

Hyphomicrobiaceae. The cells exist mainly in free-floating groups in which individual cells are attached to one another by a slender, sometimes branched filament. Multiplication is initiated by the outgrowth of a filament from the end of a mature cell or from some point on a filament connecting two mature cells. The new cell is formed by enlargement of the tip of the filament. This family consists of two genera each of which contains a single species.

The species *Hyphomicrobium vulgare* is a colorless organism that obtains its energy by the oxidation of simple organic compounds. A mature cell may develop a flagellum whose active movement can serve to tear the cell free from its filament. This organism is common in soil and fresh water and is almost always present in enrichment cultures for nitrifying bacteria. A closely related organism has been found in pus from a sinus and may be pathogenic for man.

The species *Rhodomicrobium vannieli* is pink to orange in color and is photosynthetic. The cells never develop flagella. Like *H. vulgare* it is common in fresh water and mud.

Pasteuriaceae. Members of this family are stalked bacteria which attach to surfaces. None have been isolated in pure culture and their descriptions are based mainly upon the appearance of specimens found on slides immersed in fresh water. The family contains two genera, each with a single species. In *Pasteuria ramosa* the cells are pear-shaped and the stalks are very short or lacking. It has been found in water and as a parasite of insects. In *Blastocaulis sphaerica* the cells are ical and possess long, slender stalks which arise from a common holdfast. It is fou

Hypnosis

An altered state of consciousness in which the individual is more susceptible to suggestion and in which regressive behavior may spontaneously occur Alterations in memory, perception, and sensation can readily be induced. Such changes can be incorporated into complex behavior of the individual resulting in amnesia, fuguelike conditions, paralysis, loss of sensory functions, changes in attention, and even modifications of sleeping and dreaming.

The theories of hypnosis attempt to explain its essential nature, even though they are inadequate in some respects. These theories, arranged somewhat in order of their chronological development, adduce that hypnosis is (1) a form of subtle emanation between the hypnotist and subject, (2) a modified form of sleep, (3) a form of dissociation, (4) a conditioned response resulting in ideomotor action (that is, nonvoluntary movement resulting from some idea), (5) a form of role-playing or goal-directed activity, (6) a form of regression to a level representing the relation between the parent and child, (7) a form of regression in which the subject assumes the submissive role similar to that assumed by the female in the sexual role, (8) a form of interpersonal relationship between the subject and the hypnotist, but not necessarily of the kinds represented in (6) and (7). Although the foregoing statements represent an oversimplification, they do give some idea concerning the nature of the theories.

The applications of hypnosis to the field of medicine and psychology are exceedingly wide in scope.

Hypnosis as a diagnostic tool. It has been employed successfully in distinguishing functional from organic disorders. For example, functional deafness may be distinguished from organic deafness; hysterical conversion paralysis may be differentiated from organic paralysis; psychogenic convulsions may be contrasted with organic convulsions. Hypnosis may also be used in determining a normal baseline for functioning of certain of the endocrine glands. Anxiety may bring about dysfunction of the thyroid; consequently a metabolic or other test would not distinguish between pathological disorder and a functional disorder. The anxiety can be reduced by hypnotic relaxation, and a relative relationship can be established between the output produced by an anxiety component and possible pathology.

It can also be used in teasing out the part that a functional overlay plays in what appears to be an organic condition. For example, in a limb in which there is a partial paralysis due to any organic cause, there may be superimposed a further dysfunction due to psychological causes which can be demonstrated by hypnotic suggestion.

Hypnosis as an aid in examination. It can be used in controlling pain in certain types of examination, as in cystoscopy, and leave the patient able to assist the physician by giving information concerning the location of the instrument. It has an advantage over analgesics or anesthetics in such situations. It may be useful in nerve block infiltrations and in making spinal punctures. It can also be employed successfully to reduce gagging when stomach tubes are inserted. It has been utilized in the control of bleeding, of salivation, and of muscle spasms, which are sometimes troublesome during examinations.

Hypnosis in symptom control. Hypnosis has been used effectively since the time of the British surgeon J. Esdaille (1850) for the control of pain. In cases where an analgesic or anesthetic is inadvisable, it may be used in place of, or in conjunction with, chemoagents. It may also be used where tolerance becomes a factor in long-term cases. Pain control in all kinds of surgical operations is limited only by the susceptibility of the subject. Tooth extractions, cranial operations, amputations, cardiac and lung surgery, as well as abdominal and genitourinary operations, have been performed without any kind of anesthesia except hypnosis. Intractable pain, involved in trigeminal neuralgia, phantom limb, carcinoma, burns, and other types of pain, has been allayed. Hypnosis has been used in the treatment of hiccoughs, neurodermatitis, the dumping syndrome (weakness, nausea, vertigo, and palpitation occurring immediately after partial or complete removal of the stomach), coughing spasms, headaches, frigidity, impotence, dysmenorrhea, discomfort from contact lenses, stuttering, tics, and the pain of childbirth. Many organic symptoms, as well as most functional symptoms, can be suppressed in appropriate cases. The suppression may be complete or only partially controlled, so that the patient is still able to report dangerous developments that may be occurring. The above symptoms are only a partial list of those in which hypnosis has been applied.

Hypnosis as an uncovering device. Hypnosis may be substituted for sodium amytal, sodium pentothal, or scopolamine in patient interviews to determine unconscious motivation. It will assist in recovering memories repressed because of emotional trauma. It may be an aid to surmount blocking which occurs in the process of free association. Emotional factors in functional disorders, nightmares, blackouts, and persistent dreams may be brought to light by regressive techniques.

Hypnosis as a motivational tool. This area has not been explored fully, and it merits more extensive investigation. Hypnosis has been utilized to motivate speech retraining in aphasics, to increase the desire for psychotherapy, to enhance movement or to decrease movement of limbs that are involved in surgical procedures, and to increase motivation for early ambulation following surgery. It has also been used to enhance feeling tone in some very depressed patients (*see* MEMORY; MOTIVATION; PSYCHOTHERAPY).

Hypnosis in conjunction with analysis. Hypnosis has been employed either as a form of hypno-

analysis or in conjunction with analytic procedures. In this connection either it has been used as an uncovering technique or the analysis has proceeded entirely in the hypnotic state. It has been employed particularly to overcome blocks in free association or to produce dreams. Another application is to create illusory situations through which the patient works, thus enabling the therapist to gain a better understanding of the patient's ability to cope with certain problems. Hypnosis has been utilized to enhance the feelings of schizophrenics who may develop some intellectual understanding of their

[illegible]

various pharmacological agents or may be employed with them to increase their effects; it has been used with soporifics, hypnotics, anesthetics and analgesics, tranquilizers, and with glandular, allergic, and hallucinatory agents (see PSYCHOPHARMACOLOGIC DRUGS).

It may be substituted for chlorals, barbiturates, morphine, alcohol, and other soporifics for the induction of sleep. Some of the hypnotics for which it may be used are sodium amytal, sodium pentothal, and scopolamine. In producing anesthesia or analgesia it has been substituted for, or used in conjunction with, chloroform, nitrous oxide, procaine, and other similar agents. Relaxation and anxiety, which are ordinarily controlled by Thorazine, Frenquil, Serpasil, Equanil, and other tranquilizers, can be controlled by hypnosis in suitable patients. In

...duction of transitory hallucinatory phenom-
...ychosis, ... rimental purposes, pe-
...ysergi... ten employed. Similar
...e cre... mosis.
...ha... nts mentioned have
...nts... lopment of toler-
...ec... plications. Hyp-
...o... ve these disad-
... mended
... rs in-
... ise

nervous system depressants but their hypnotic effects, duration of action, margin of safety, and habit-forming properties vary considerably. Most are synthetic ureides, that is, substitution products of urea and an organic acid. They are usually administered in capsule form as well as by other routes. See BARBITURATES; [illegible].

Bromine compounds comprise another group of hypnotics whose effect is achieved by depression of central nervous system activity by the bromide ion. Although popular and overexploited, the bromides usually exert an effect only after multiple doses spread over a period of time, hence [illegible] their relative effectiveness as anticonvulsant agents. Chronic bromine intoxication is not uncommon after overuse.

Chloral hydrate is one of the oldest, cheapest, and best hypnotics in use if properly prescribed. It is a chlorinated derivative of ethyl alcohol that produces a more nearly physiological sleep.

Paraldehyde, a polymer of acetaldehyde, is one of the best hypnotics with a low toxicity. Because of its unpleasant odor and taste it is used less now than formerly.

Habituation is a problem in the psychologically susceptible. [illegible]

Hypochlorite

The negative ion, ClO^-, derived from hypochlorous acid, HClO. The hypochlorite ion is best known as an oxidizing agent and as a constituent of bleaching agents. Hypochlorites are prepared by passing chlorine gas into alkaline solution whereupon the chlorine is simultaneously oxidized and reduced

$$Cl_2 + 2NaOH \rightarrow NaCl + NaClO + H_2O$$

Sodium hypochlorite is extensively used as a bleach in the textile industry and in laundries and as a disinfectant in the home.

When chlorine is passed over slaked lime, a product called bleaching powder or chlorinated lime is obtained which approximates the formula CaCl(OCl). A related preparation, High Test Hypochlorite (H.T.H.), which contains twice as much available chlorine as bleaching powder, may be represented by the formula $Ca(OCl)_2$.

These preparations are shipped in the powder ...rm to the point of use and converted to sodium ...ochlorite by the addition of soda lime as foll... ...ds:

$$Na_2CO_3 + Ca(ClO)_2 \rightarrow 2NaClO + CaCO_3$$

ANTIMICROBIAL AGENTS; BLEACHING; CHLO... ; CHLORINE; OXIDIZING AGENT; PERCHLORATE. [E.E.W.]

...dermis

...podermis, or exodermis, is the outermost ...r of the cortex. It forms a prominent layer ...ely under the epidermis in many but not ... Like the endodermis, it develops Casparian ..., suberin deposits, and cellulose deposits ...ted with phenolic or quinoidal substances

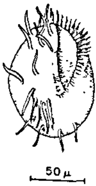
Hypothesis

A tentative supposition with regard to an unknown state of affairs, the truth of which is thereupon subject to investigation by any available method, either by logical deduction of consequences which may be checked against what is known, or by direct experimental investigation or discovery of facts not hitherto known and suggested by the hypothesis. A hypothesis differs from a postulate in that it is meaningful to inquire as to the objective truth of a hypothesis, and the chief interest in framing a hypothesis is to find one which is true. Often it is possible to estimate the truth of a hypothesis only on a probability basis.

A classical example is the nebular hypothesis with regard to the origin of the solar system. The "truth" of this hypothesis may be checked by its ability to explain the properties of the solar system as now observed [P.W.B.]

Hypotrichida

A well-known order of Spirotricha. These animals are commonly considered to represent the pinnacle of specialized development in the evolution of ciliates. Somatic ciliature, of the ventral surface, has been replaced by cirri; it is absent or represented by inconspicuous sensory bristles on the dorsal surface. The adoral zone of membranelles is very



Hypotrichida. (a) *Euplotes*. (b) *Diophrys*.

prominently developed and the buccal area may occupy a large part of the ventral surface of the body. The whole body is generally rigid and is pronouncedly flattened dorsoventrally. Hypotrichs occur ubiquitously in fresh- and salt-water habitats and have been studied quite extensively by experimental biologists. Common examples are *Euplotes* and *Diophrys* (see illustration), as well as *Oxytricha* and *Stylonychia*. *See* SPIROTRICHA. [J.O.C.]

Hyracoidea

An order of small

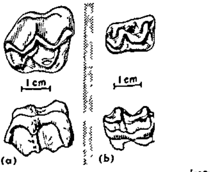
fossil genera are known. The hyraxes resemble rabbits in appearance and general structure, and were long thought to be rabbits of some sort. They are now known to be archaic hoofed animals. The incisor teeth are rootless and rodentlike, separated by a gap from the cheek teeth, which are high-crowned and with molarized premolars. There are four functional toes in front and three behind, each provided with a hooflike nail. Hyraxes live among rocks or in trees, and all are excellent climbers. The tree hyraxes (*Dendrohyrax*) are the only truly arboreal ungulates. *See* HYRACOIDEA FOSSILS; EUTHERIA. [D.D.D.]

Hyracoidea fossils

Hyracoids or coneys are a group of quadrupedal, plantigrade, herbivorous, ungulate mammals well diversified and widespread in the Oligocene and Miocene deposits of Africa. A few specimens have been recovered from the middle and late Pliocene of Greece and Southern France and are reported from the Pleistocene of China. The group was probably more widely distributed during Pliocene time in Eurasia, but not as an abundant element of the faunas.

The order is separated into three families: the Procaviidae or short-snouted, moderate-sized forms with the longest geologic history and leading to the living representatives; the Geniohyidae or long-snouted, giant forms of the Oligocene and Miocene that reached the size of a tapir; and the Myohyracidae or pygmy forms of the Miocene with rodentlike adaptations in the incisor teeth.

Although the hyracoids have been compared with members of the Notoungulata, Proboscidea, Litopterna, and others, recently they have been found to possess many affinities with the Perissodactyla. These affinities are to be found in the dentition, skull, and appendicular skeleton. They are probably descended from some Paleocene condylarthlike stock of the New World that gave rise to the Meniscotheriidae and Phenacodontidae presently in the order Condylarthra, as well as the Perissodactyla and Artiodactyla. *See* ARTIODACTYLA FOSSILS; CONDYLARTHA; PERISSODACTYLA FOSSILS.



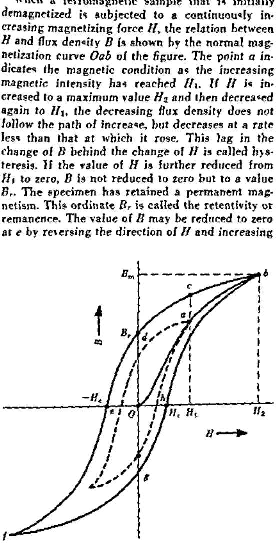
(a) Upper and (b) lower molar of *Megalohyrax palaeotherioides* from the early Oligocene of Egypt (After M. Schlosser, 1911)

The evolutionary trend in the order has been primarily one of reduction in size, loss of teeth, and shortening of snout. Oligocene and Miocene genera have a closed (no large spaces between anterior teeth) and complete dentition; some have long and others short snouts; and all except the two genera of the Myohyracidae are large forms. The small rabbit-sized later forms with short snouts show a progressive loss of lateral incisors and canines. All have an elephantlike adaptation in the feet in which the toes are drawn together with a pad beneath. Although a well diversified and abundant group in the middle Tertiary, they have dwindled to three living genera restricted to Africa and Syria. [C.T.J.]

Hysteresis, magnetic

The lagging of magnetization behind magnetizing force as the magnetic condition of a ferromagnetic material is changed.

When a ferromagnetic sample that is initially demagnetized is subjected to a continuously increasing magnetizing force H, the relation between H and flux density B is shown by the normal magnetization curve *Oab* of the figure. The point a indicates the magnetic condition as the increasing magnetic intensity has reached H_1. If H is increased to a maximum value H_2 and then decreased again to H_1, the decreasing flux density does not follow the path of increase, but decreases at a rate less than that at which it rose. This lag in the change of B behind the change of H is called hysteresis. If the value of H is further reduced from H_1 to zero, B is not reduced to zero but to a value B_r. The specimen has retained a permanent magnetism. This ordinate B_r is called the retentivity or remanence. The value of B may be reduced to zero at e by reversing the direction of H and increasing



Hysteresis loop.

its value to H_c. This value H_c is called the coercive force or coercivity.

Hysteresis loop. As H is increased in the negative direction, the magnetization proceeds along the curve of the figure until at f the values of B and H are the same as those at b, but opposite in direction. When the reverse changes in H are made, the magnetization changes along the curve *fghb*. This entire loop *bdefghb* is called a hysteresis loop. If the hysteresis loop starts from another point on the normal magnetization curve, such as a, there will be a smaller hysteresis loop entirely within the larger, such as the dotted loop of the figure. For additional information on magnetization curves, see MAGNETIZATION.

Energy. In magnetizing the core, energy must be supplied. As H and B increase along *fghb*, the energy gain is proportional to the area under that portion of the curve. Along the path *bcdef*, there is a loss in energy proportional to the area under *bcdef*. The net loss in energy per cycle per unit volume is

$$W = \oint dW = \oint H\,dB$$

where the integral is taken around the closed loop. But $\oint H\,dB$ is the area of the hysteresis loop, and the energy loss per unit volume per cycle is equal to the area of the hysteresis loop. This energy is converted into heat. If B is expressed in webers/m^2 and H in ampere turns/m, the energy loss is in joules/(m^3) (cycle).

C. P. Steinmetz found an empirical relation between the energy loss per unit volume per cycle W and the maximum value B_m of the flux density during the cycle

$$W = \eta(B_m)^n$$

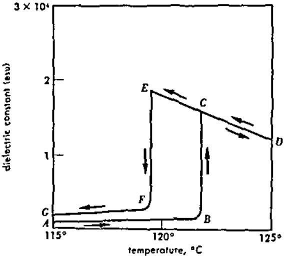
where η is called the Steinmetz coefficient. Steinmetz found a value of about 1.6 for the exponent n for many materials, but it varies from about 1.5 to 2.5 for others. Values of η have been tabulated.

In alternating-current machinery, masses of iron are in fields that are constantly reversing. Therefore, the iron is constantly being carried around hysteresis paths, and there is an energy loss per cycle that depends upon the hysteresis loop for the particular iron that is used. This hysteresis loss results in undesired heating of the iron as well as waste of energy. For additional information on hysteresis loss, *see* CORE LOSS; *see also* HYSTERESIS MOTOR; HYSTERESIS, THERMAL. [K.V.M.]

Bibliography: E. R. Peck, *Electricity and Magnetism*, 1953; F. W. Sears, *Principles of Physics*, vol 2, 1951; R. P. Winch, *Electricity and Magnetism*, 1955.

Hysteresis, thermal

A phenomenon sometimes observed in the behavior of a temperature-dependent property of a body. Thermal hysteresis is said to occur if the behavior of such a property is different when the body is heated through a given temperature range fro



Plot of dielectric constant vs. temperature for a single crystal of barium titanate (*After M. E. Drougard and D. R Young, Phys. Rev., 95:1152–1153, 1954*)

when it is cooled through the same temperature range The accompanying figure shows the thermal hysteresis which has been observed in the behavior of the dielectric constant of single crystals of barium titanate in the neighborhood of 120°C. On heating, the dielectric constant was observed to follow the path *ABCD*, and on cooling the path *DCEFG*. [J.D.L.]

Hysteresis motor

A synchronous motor without salient poles and without dc excitation which makes use of the hysteresis and eddy-current losses induced in its hardened-steel rotor to produce rotor torque. The stator and stator windings are similar to those of an induction motor and may be polyphase, shaded-pole, or capacitor type. The rotor is usually made up of a number of hardened steel rings on a nonmagnetic arbor. The hysteresis motor develops constant torque up to synchronous speed. The motor can, therefore, synchronize any load it can accelerate. These motors are built in small sizes only, such as those for electric clocks and phonographs. *See* SYNCHRONOUS MOTOR; *see also* INDUCTION MOTOR. [L.V.B.]

Hysteria

A type of neurosis. It was with the problem of hysteria that Sigmund Freud began his critical observations on the neurotic process. He noted the development of sensory anesthesias and motor paralyses which had no neurological origin but which seemed to express a way of defending against unacceptable unconscious wishes. An example is the soldier who cannot admit his fear of facing battle and develops a nonmalingering motor paralysis without neurological base (*see* NEUROSIS). Hysterical symptoms, however, are not always somatic in nature. What appears universal about them is the presence of massive repression and the development of a symptom pattern that indirectly or symbolically expresses the repressed needs and wishes. An example is the fugue state or, of the same form but less severe, the amnesias, both of which represent attempts to flee from the stressful threat with which the person is faced. Another symptom of hysterical reaction is the impoverished affect of the person when repression of impulses related to the core difficulty has been carried so far as to produce an impoverished personality. The hysterical case usually exhibits a high degree of suggestibility, particularly to the physician, and this suggestibility, along with the tendency for symptoms to be split off from the rest of the person's functioning, led P. Janet and others to conceive of hysteria as a disorder based on dissociation of aspects of the personality. *See* ABNORMAL BEHAVIOR. [J.S.B.; W.M.S.]